THE PROCEEDINGS OF
THE PHYSICAL SOCIETY
Section B

VOL. 70, PART 1 1 January 1957 No. 445 B

Thermonuclear Reaction Rates

BY W. B. THOMPSON
Atomic Energy Research Establishment, Harwell, Didcot, Berks.

Communicated by B. H. Flowers; MS. received 12th November 1956

Abstract. It is shown that the thermonuclear reaction rate between two gases in thermal equilibrium and of densities n_1 and n_2 nuclei per unit volume can be written as $n_1n_2\langle\sigma v\rangle$ where $\langle\sigma v\rangle$ is a suitable average of the product of the reaction cross section and relative velocity, and for any reaction is a function of temperature alone. These functions are given for a number of exoergic reactions between light nuclei.

§ 1. INTRODUCTION

As a preliminary to any study of the feasibility of obtaining power from controlled thermonuclear reactions it is necessary to estimate the rate of production of energy, and hence the reaction rate, as a function of temperature. This paper gives the reaction rates for a number of processes to as high a degree of accuracy as is compatible with our limited knowledge of the low-energy cross sections, and describes the calculation of these rates. The particular reactions considered, all exoergic and all involving deuterium or tritium, are listed in table 1 together with some information about the cross sections and energy release.

Table 1.

Reaction	(1)	(2)	(3)
^{2}H(d, p)^{3}H	4·03	0·079	1
^{2}H(d, n)^{3}He	3·25	0·09	1
^{3}H(d, n)^{4}He	17·58	5·0	0·1
^{3}He(d, p)^{4}He	18·34	0·8	0·45
^{3}H(t, 2n)^{4}He	11·32	0·098	1
^{6}Li(d, p)^{7}Li	5·02	0·01	1
^{6}Li(d, n)^{7}Be	3·37 } total	0·016	1
^{6}Li(d, n)^{4}He+^{3}He	1·79 }		
^{6}Li(d, α)^{4}He	22·36	0·026	0·6
^{7}Li(d, n)^{8}Be	15·01	0·045	1
^{7}Li(d, n)2^{4}He	15·1	—	—

(1) Energy released (MEV); (2) maximum cross sections below 1 MEV (barns); (3) bombarding energy for maximum cross section (MEV).

The reaction rate r is equal to the product of the number of possible pairs of reacting particles and the probability of a reaction between any pair, i.e. $r = n_1n_2\langle\sigma v\rangle$ for reactions between unlike particles and $r = \frac{1}{2}n^2\langle\sigma v\rangle$ for reactions between like particles. Here n_1, n_2 and n are the number densities of reacting particles, and the reaction probability is written in the significant form $\langle \sigma v \rangle$ to indicate that it is an appropriate average of the product of the reaction cross section and relative velocity of the interacting particles. If the reacting mixture is in thermal equilibrium this quantity depends only on the temperature and as shown in § 2 can be written as a single integral and approximated in certain cases by an explicit function.

The cross sections used have been taken from a compilation by Hirst (1953) of published data concerning exoergic reactions between light nuclei. The three most important cross sections, those for reactions D + D, D + T and D + ^{3}He have recently been re-examined by Arnold *et al.* (1956) and for D + D and D + ^{3}He their results are in agreement with those given by Hirst, but for D + T they report cross sections substantially higher at low energies than those given by Hirst. These correspond to a shift in the energy scale of about 6% at 30 kev, which is considerably greater than their estimated maximum error. For this reaction the reaction rates have been calculated using both Hirst's curve and the Gamow function fitted to the data of Arnold *et al.*

§ 2. THE REACTION PROBABILITY

The rate r at which a reaction proceeds in a uniform gas mixture may be written in terms of the cross section σ and the velocity distribution functions $n_1(v)$ and $n_2(v)$ of the components in the form

$$r = \int\int n_1(\mathbf{v}_1) n_2(\mathbf{v}_2) |\mathbf{v}_1 - \mathbf{v}_2| \sigma(\mathbf{v}_1 - \mathbf{v}_2) d^3\mathbf{v}_1 d^3\mathbf{v}_2 \qquad \ldots\ldots(2.1)$$

where the volume integrals cover both velocity spaces.

If the components are in thermal equilibrium at temperature T the distributions are Maxwellian; hence

$$r = n_1 n_2 \left(\frac{m_1}{2\pi kT}\right)^{3/2} \left(\frac{m_2}{2\pi kT}\right)^{3/2} \int\int d^3\mathbf{v}_1 d^3\mathbf{v}_2 |\mathbf{v}_1 - \mathbf{v}_2| \sigma(\mathbf{v}_1 - \mathbf{v}_2) \times \exp\left\{-\frac{\frac{1}{2}m_1 v_1^2 + \frac{1}{2}m_2 v_2^2}{kT}\right\} \qquad \ldots\ldots.(2.2)$$

where n_1, n_2, m_1, m_2 are the number densities and particle masses of the two particles, and k is Boltzmann's constant.

The integral may be reduced by changing to the relative velocity $\mathbf{v}$ and the velocity $\mathbf{V}$ of the centre of mass, where $\mathbf{v} = \mathbf{v}_1 - \mathbf{v}_2$ and $\mathbf{V} = (m_1\mathbf{v}_1 + m_2\mathbf{v}_2)/(m_1 + m_2)$, whereupon the integral over $d^3\mathbf{V}$ and over the solid angle in $d^3\mathbf{v}$ may be carried out immediately. We obtain

$$r = n_1 n_2 \, 4\pi \left(\frac{m_1 m_2}{2\pi kT(m_1 + m_2)}\right)^{3/2} \int_0^\infty dv \, v^3 \sigma(v) \exp\left\{-\tfrac{1}{2}\frac{m_1 m_2}{m_1 + m_2}\frac{v^2}{kT}\right\}, \quad (2.3)$$

This is most naturally expressed in terms of the relative energy; but since the cross sections are usually given as a function of the energy of the bombarding particle $\frac{1}{2}m_1 v^2$ it is most convenient for computation to use this energy as the variable of integration. If we introduce instead of T the value W of kT in units of kev, so that $T(°\text{K}) = 11{\cdot}6 \times 10^6 W$, the reaction probability becomes

$$\langle \sigma v \rangle = 4{\cdot}94 \times 10^{-17} \left(\frac{m}{m_1}\right)^{3/2} \left(\frac{W}{m_1}\right)^{1/2} W^{-2} \int_0^\infty dE \, E\sigma(E) \exp\left\{-\frac{m}{m_1}\frac{E}{kT}\right\} \text{cm}^3\,\text{sec}^{-1} \qquad \ldots\ldots(2.4)$$

where m_1 and $m(=m_1m_2/(m_1+m_2))$ are the mass of the first particle and the relative mass in atomic units, while $\sigma(E)$ is the cross sections in barns ($=10^{-24}\,\mathrm{cm}^2$) as a function of the energy of the first particle.

This quantity has been computed under the direction of Mr. E. B. Fossey and the results are given in table 2 and in figures 1 and 2.

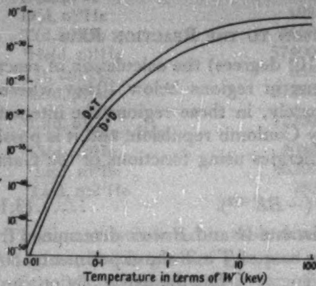


Figure 1.

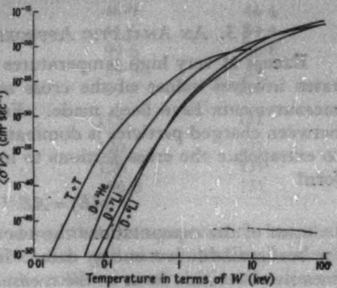


Figure 2.

Figure 1. Reaction probabilities for D+D and D+T. $\langle\sigma v\rangle$ plotted against temperature (in terms of *W*).

Figure 2. Reaction probabilities for T+T, D+³He, D+⁶Li and D+⁷Li. $\langle\sigma v\rangle$ plotted against temperature (in terms of *W*).

Table 2. Reaction Probabilities $\langle\sigma v\rangle$ (cm³ sec⁻¹)

Temp. (°K) *W* (kev)	10^5 0·01	10^6 0·1	10^7 1	10^8 10	10^9 100
$D(d, p)^3H$	2·61 (52)	6·69 (32)	8·78 (23)	6·12 (19)	2·34 (17)
$D(d, n)^3He$	2·61 (52)	6·69 (32)	8·78 (23)	5·71 (19)	2·67 (17)
D+D (total)	5·22 (52)	1·34 (31)	1·76 (22)	1·18 (18)	5·01 (17)
D+D (app.)	2·5 (52)	0·87 (31)	1·57 (22)	1·25 (18)	4·0 (17)
$T(d, n)^4He$	6·1 (50)	6·22 (30)	5·56 (21)	10·5 (17)	8·0 (16)
T+D (app.)	10·0 (50)	8·0 (30)	5·7 (21)	3·1 (17)	8·0 (16)
T+D (A)	1·2 (50)	5·0 (30)	9·3 (21)	8·5 (17)	2·6 (16)
$^3He(d, p)^4He$	2·3 (79)	4·73 (43)	1·49 (26)	7·43 (20)	4·52 (16)
$T(t, 2n)^4He$	2·51 (56)	1·64 (34)	1·41 (24)	9·8 (21)	—
$^6Li(d, p)^7Li$	1·02 (86)	1·13 (47)	6·5 (30)	1·8 (20)	1·3 (16)
$^6Li(d, n)\genfrac{}{}{0pt}{}{^7Be}{^3He+{}^4He}$	6·05 (93)	1·13 (49)	6·06 (30)	3·89 (21)	1·93 (17)
$^6Li(d, \alpha)^4He$	1·02 (95)	3·31 (51)	6·68 (31)	5·26 (22)	4·9 (18)
$^7Li(d, n)^8Be$	7·36 (100)	1·13 (50)	6·74 (32)	1·03 (21)	2·79 (17)
$^7Li(d, n)2^4He$	1·88 (86)	1·54 (47)	7·8 (30)	1·27 (21)	—

Notes : The number in parenthesis indicates the negative power of 10 to be associated with the first factor : thus 2·61 (52)$=2{\cdot}61\times10^{-52}$. T+D (A) has been computed from the cross sections given by Arnold *et al.*

From the results it is immediately obvious that the most hopeful reaction for the production of thermonuclear energy is the T+D which has a reaction probability at least ten times greater than any other for temperatures between 10^5 10^9 degrees. The D+D reaction although proceeding more slowly uses only naturally occurring material, and if all the charged reaction products can be made to take part in the thermonuclear process the total energy released per reaction is 21·6 Mev, slightly greater than the 17·6 Mev produced by T + D alone. At temperatures of less than 10^8 degrees only these reactions are significant, although this may not hold at higher temperatures.

§ 3. An Analytic Approximation to the Reaction Rate

Except at very high temperatures ($> 10^8$ degrees) the calculation of reaction rates involves values of the cross sections in regions below 10 kev where no measurements have been made. Fortunately, in these regions the interaction between charged particles is dominated by Coulomb repulsion, and it is possible to extrapolate the cross sections to low energies using functions of the Gamow form

$$\sigma(E) = AE^{-1}\exp(-BE^{-1/2}). \qquad \ldots\ldots(3.1)$$

In most of the computations the two constants A and B were determined from the best available low energy data. In the case of T + T the experimental curve does not have this form at the measured energies, but the extrapolation can be made using the theoretical value of B,

$$BE^{-1/2} = 2\pi Z_1 Z_2 \frac{e^2}{\hbar v} = \pi\sqrt{2}\alpha(m)^{1/2} c \frac{Z_1 Z_2}{\frac{1}{2}mv^2} = 31{\cdot}4 Z_1 Z_2 (m)^{1/2} E^{-1/2} \qquad \ldots\ldots(3.2)$$

where m is the relative mass in atomic units, E the relative energy in kev and Z_1 and Z_2 the charges of the two particles.

When (3.1) is an adequate representation of the cross section, it is possible to get a reasonable analytic approximation to the reaction probability. This is most easily obtained by introducing (3.1) into (2.3) which becomes

$$r = n_1 n_2\, 2\left(\frac{2}{\pi}\right)^{1/2}\left(\frac{kT}{m}\right)^{1/2}\frac{A}{kT}\frac{1}{kT}\int_0^\infty dE\exp\left\{-\left(\frac{E}{kT}+BE^{-1/2}\right)\right\}. \qquad \ldots\ldots(3.3)$$

The integral $\int_0^\infty dx\ \exp\{-(x + ax^{-1/2})\}$

may be approximated by $2(\frac{1}{3}\pi)^{1/2}(\frac{1}{2}a)^{1/3}\exp\{-3(\frac{1}{2}a)^{2/3}\}$ and then

$$\begin{aligned}\langle\sigma v\rangle &= 4\left(\frac{2}{27}\right)^{1/6}\left(\frac{kT}{m}\right)^{1/2}\frac{A}{kT}\left(\frac{B^2}{4kT}\right)^{1/6}\exp\left\{-3\left(\frac{B^2}{4kT}\right)^{1/3}\right\}\\ &= 0{\cdot}8\times 10^{-16}AB^{1/3}m^{-1/2}W^{-2/3}\exp\{-3(\tfrac{1}{4}B^2)^{1/3}W^{-1/3}\}\\ &\qquad\qquad \text{cm}^3\ \text{sec}^{-1}, \qquad \ldots\ldots(3.4)\end{aligned}$$

where A is in barns kev and B in $(\text{kev})^{1/2}$.

To use this formula the cross section must be expressed as a function of the relative energy. Table 3 contains the values of the constants A', B' giving the cross section as a function of the energy of the bombarding particle. The relation between the constants is $A = mA'/m_1$, $B = B'\ (m/m_1)^{1/2}$. Reaction probabilities computed by this approximate formula are included in table 2. It will be noted that they differ from those computed directly by factors usually less than 2, hence are adequate for all order of magnitude calculations.

A further result follows from the method of obtaining the analytic approximation—a determination of that part of the distribution function contributing the major part of the reaction rate. This is the region near $E_m = (BW/2)^{2/3}$.

Table 3. Gamow Constants used to fit Cross Sections as function of Bombarding Energy E in the form $\sigma(E) = A'E^{-1} \exp(-B'E^{-1/2})$

	A' (barns kev)	B' (kev$^{1/2}$)	B'_{theor}
D(d, p)^{3}H	198	46·9	44·4
D(d, n)^{3}He	198	46·9	44·4
T(d, n)^{4}He	7660	41·6	44·4
^{3}He(d, p)^{4}He	57400	99·5	88·8
T(t, 2n)^{4}He	668	59·9	54
^{6}Li(d, p)^{7}Li	27200	99·0	131
$^6\mathrm{Li}(d, n)^{^7\mathrm{Be}}_{^3\mathrm{He}+{}^4\mathrm{He}}$	12200	115	131
^{6}Li(d, α)^{4}He	4080	120	131
^{7}Li(d, n)^{8}Be	46200	139	131
^{7}Li(d, n)2^{4}He	283	96·8	131

This quantity is listed in table 4 which shows that for temperatures less than 10^8 degrees, the important group of particles lies well out on the high-energy tail of the Maxwellian distribution. Since the collision cross section for charged particles is small at high energies a Maxwellian distribution is only slowly established, the high energy particles can easily be lost at moderate temperatures, and hence the calculated reaction rates might be difficult to achieve except at the highest temperatures.

Table 4. Relative Energy (kev) of Particles making Maximum Contribution to Reaction

Temp. (°K)	10^5	10^6	10^7	10^8	10^9
W(kev)	0·01	0·1	1·0	10	100
D+D	0·29	1·35	6·2	29	135
D+T	0·31	1·42	6·6	31	140
D+^{3}He	0·49	2·3	11·0	49	230
T+T	0·33	1·5	7·2	33	155
D+^{6}Li	0·68	3·2	15	68	320
D+^{7}Li	0·68	3·2	15	68	320

Some care is needed in using this table. For temperatures less than 10^8 degrees the most important collisions are head-on and the laboratory energy of the significant particles is a good deal less than the relative energy. For the reactions listed and $T < 10^8$, the important deuteron energy is about one-half the relative energy.

REFERENCES

ARNOLD, W. R., PHILLIPS, J. A., SAWYER, G. A., STOVALL, E. J., and TUCK, J. L., 1956, *Phys. Rev.*, 93, 483.

HIRST, F., 1953, Atomic Energy Research Establishment, Report N/M 64.

Some Criteria for a Power Producing Thermonuclear Reactor

By J. D. LAWSON

Atomic Energy Research Establishment, Harwell, Berks.

Communicated by D. W. Fry; MS. received 2nd November 1956

Abstract. Calculations of the power balance in thermonuclear reactors operating under various idealized conditions are given. Two classes of reactor are considered: first, self-sustaining systems in which the charged reaction products are trapped and, secondly, pulsed systems in which all the reaction products escape so that energy must be supplied continuously during the pulse. It is found that not only must the temperature be sufficiently high, but also the reaction must be sustained long enough for a definite fraction of the fuel to be burnt.

---

§ 1. Introduction

It has been widely conjectured that some form of controlled thermonuclear reactor, capable of producing a useful amount of power, will some day be constructed. In this paper the power balance in such a reactor is considered, and some criteria which have to be satisfied in a power producing system are derived.

Some of the difficulties of realizing a controlled fusion reaction have been discussed by Thirring (1955) and Thonemann (1956), and a broad survey of fundamentals has been given by Post (1956). The present treatment differs from that of Thirring in the assumptions about the radiation from a hot gas, and it covers in rather more detail some of the points discussed by Thonemann and Post. The analysis is based on simple assumptions; it is designed to illustrate the essential features of the problem, and is neither rigorous nor complete. The assumptions made are in all cases optimistic, so that the criteria established are certainly necessary, though by no means sufficient, for the successful operation of a thermonuclear reactor.

§ 2. Basic Principles

Of the exoergic reactions involving light nuclei those between the hydrogen isotopes (the so-called D–D and T–D reactions) are by far the most probable at low energies. Of these the T–D reaction has the higher cross section, but since tritium does not occur naturally it is necessary to use a system in which it can be bred. This may be done by capturing the neutrons emitted in the T–D reaction in ^{6}Li, which then decays into T and ^{4}He.

The reactions of interest are shown in table 1.

Table 1

Reaction	Q (Mev)	$\sigma_{\max}$ (barns)	$\sigma_{\max}$ energy
^{2}H(d, n)^{3}He	3·3	0·09	2 Mev
^{2}H(d, p)^{3}H	4·0	0·16	2 Mev
^{2}H(t, n)^{4}He	17·6	5·0	150 kev
^{6}Li(n, α)^{3}H	4·8	$1/v$ law	

The energy released per unit time and volume by thermonuclear reactions in a hot gas is given by

$$P_R = n_1 n_2 \overline{v\sigma}(T) E \qquad \ldots\ldots(1)$$

where n_1 and n_2 are the number densities of the nuclei of the first and second kinds, and $\overline{v\sigma}(T)$ is the product of the relative velocities of the nuclei and the reaction cross section averaged over the Maxwellian velocity distribution corresponding to a temperature T, and E is the energy released by one reaction. If the ions are of the same kind (as in the D–D reaction) $n_1 n_2$ is replaced by $2(n/2)^2 = \frac{1}{2}n^2$. Values of $\overline{v\sigma}(T)$ calculated from published values of the cross sections for the D–D and T–D reactions are given in the companion paper by Thompson (1957).

Energy can be lost from the hot gas in two ways, by radiation and by conduction. At temperatures above about 10^6 degrees hydrogen is completely ionized and radiation occurs principally as bremsstrahlung (free–free transitions). The mean free path of such radiation is large (several g cm^{-2}) and consequently in a reactor of controllable size virtually all of it would escape. The Stefan–Boltzmann T^4 law does not hold under these circumstances; the variation of intensity with temperature can only be found by a detailed study of the radiation process. The power radiated per unit volume in hydrogen is given by (Spitzer 1956)

$$P_B = 1{\cdot}4 \times 10^{-34} n^2 T^{1/2} \text{ watts cm}^{-3}. \qquad \ldots\ldots(2)$$

If the hot gas is in a magnetic field the electrons will move in spiral orbits, and additional radiation due to the acceleration towards the axis of the spiral will occur. This radiation is similar to that obtained from electrons in a betatron, and it may be important in very intense fields. It will, however, be neglected in this paper.

Conduction loss is difficult to treat in a general way, since it depends on the geometry of the system, its density and temperature distribution, and also the wall material. In the analysis which follows it is optimistically assumed that the conduction loss is zero.

It is of interest to see at what temperature the nuclear power release is equal to the radiated power. This may be called the 'critical temperature' and is the hypothetical temperature which would be needed for a self-sustaining system if all the radiation escaped but the reaction products were retained. The critical temperature is about 150 million degrees for the D–D reaction (assuming that the tritium is burnt as soon as it is formed, but that the ^{3}He is not burnt), and 30 million degrees for the T–D reaction.

The critical temperature is a somewhat artificial concept; it does not mean that if a thermonuclear fuel is heated to this temperature a reaction will be set off in the way that a chemical explosion is set off, or that the fuel can be ignited as in a gas jet. This would only be true if the energy of the reaction were deposited close to where it was released, i.e. if the range of the reaction products were short compared with the dimensions of the apparatus. In fact, the range of the particles will almost certainly be large compared with the dimensions of the apparatus if the system is to be of controllable size, so that unless the tracks are somehow coiled up it will be the walls of the apparatus which are heated rather than the gas, and energy must be fed in continuously to sustain the reaction.

Various types of system will now be considered in a general way. No suggestions of how to realize them will be given.

§ 3. SYSTEMS IN WHICH THE REACTION PRODUCTS ARE RETAINED

It is not inconceivable that the charged reaction products could be contained in the hot gas by a suitable combination of electric and magnetic fields, though it seems unlikely that the escape of neutrons can be prevented. The temperature at which such a system would be self-sustaining in the absence of conduction loss can be calculated by equating the radiation loss to the energy carried by the charged disintegration products. This temperature is about 3×10^8 degrees for the D–D reaction, and 5×10^7 degrees for the T–D reaction. In the D–D system it is only just possible in principle to sustain the reaction, since above about 10^8 degrees the reaction rate increases with temperature only slightly faster than the radiation loss. At 10^9 degrees for example a conduction loss equal to the radiation loss is sufficient to quench the reaction.

As an example of the orders of magnitude involved, the slowing down range of the charged reaction products in a gas at 10^8 degrees and 10^4 atmospheres pressure ($n = 3 \times 10^{17}$ nuclei/cm^3) is of the order of a kilometre. The range of the neutrons is hundreds of kilometres.

§ 4. SYSTEMS IN WHICH THE REACTION PRODUCTS ESCAPE

An alternative type of system in which the reaction products are not retained in the gas will now be considered. Since some specific proposals are for pulsed systems we shall consider the following idealized cycle: the gas is heated instantaneously to a temperature T, this temperature is maintained for a time t, after which the gas is allowed to cool. Conduction loss is neglected entirely, and it is assumed that the energy used to heat the gas and supply the radiation loss is regained as useful heat.

An important parameter R will now be introduced; this is the ratio of the energy released in the hot gas to the energy supplied. Now the energy released by the reaction appears as heat generated in the walls of the apparatus, and this has to be converted to electrical, mechanical or chemical energy before it can be fed back into the gas. If η is the efficiency with which this can be done, then the condition for a system with a net power gain is

$$\eta(R+1) > 1. \qquad \ldots\ldots(3)$$

The maximum value of η is about $\frac{1}{3}$, so that R must be greater than 2.

For the pulsed cycle described above we have

$$R = \frac{tP_{\mathrm{R}}}{tP_{\mathrm{B}} + 3n\boldsymbol{k}T} = \frac{P_{\mathrm{R}}/3n^2\boldsymbol{k}T}{P_{\mathrm{B}}/3n^2\boldsymbol{k}T + 1/nt} \qquad \ldots\ldots(4)$$

where P_{R} and P_{B} are respectively the reaction power and radiated power per unit volume. The $3n\boldsymbol{k}T$ term represents the energy required to heat the gas to a temperature T. Electron binding energies are neglected, but the contribution from electrons is included (this accounts for the factor 3 rather than $\frac{3}{2}$).

Since P_{R} and P_{B} are both proportional to n^2, R is a function of T and nt. In figure 1 curves of R against T for various values of nt are shown for the D–D reaction assuming that the tritium formed is also burnt. (In practice the tritium would have to be collected and fed back into the system with the deuterium.) The line $R = 2$ is shown dotted in the figure, and it is seen that for a useful reactor T must exceed 2×10^8 degrees and nt must exceed about 10^{16}. Thus, for a pulse of 1 microsecond duration, n must be greater than 10^{22}; this corresponds to

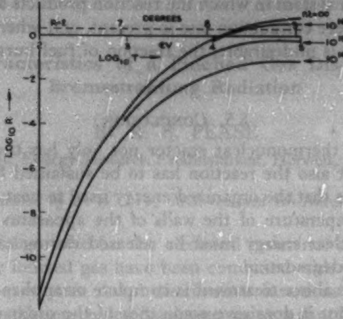


Figure 1. Variation of *R* with *T* for various values of *nt* for D–D reaction.

a pressure of 6×10^8 atmospheres at a temperature of 2×10^8 degrees. Particles move several metres during this time, and the mean free path for momentum transfer is several centimetres even at this high density. These distances are, of course, measured along the track, which may be spiralled or oscillatory.

Figure 2 shows similar curves for the T–D reaction. Conditions are easier, but still severe; *nt* must exceed 10^{14}, and the minimum temperature is 3×10^7 degrees.

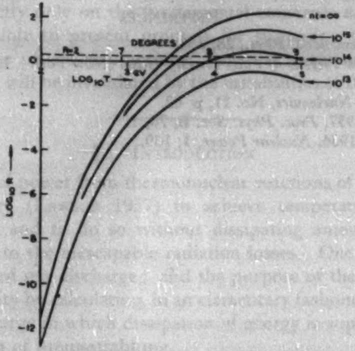


Figure 2. Variation of *R* with *T* for various values of *nt* for T–D reaction.

The curve marked $nt = \infty$ merely shows the ratio of the thermonuclear power release to the radiation loss, and crosses the $\log R = 0$ axis at the critical temperature. The curves are only accurate so long as t is sufficiently short that only a small fraction of the fuel is burnt. The values $nt = 10^{16}$ for a D–D system and 10^{14} for a T–D system both correspond to a burning of about 1% of the fuel.

Although these calculations refer specifically to a system in which the reaction products escape, it may easily be verified that the '1% burn-up' criterion is almost unaltered in a system in which the reaction products are retained. In any practical system, where conduction loss is present, and where a large circulating power in the system is undesirable, the fraction of fuel burnt would need to be much greater.

§ 5. CONCLUSION

For a successful thermonuclear reactor not only has the temperature to be sufficiently high, but also the reaction has to be sustained for a sufficient time. The reason for this is that the organized energy used to heat the gas is ultimately degraded to the temperature of the walls of the apparatus and, consequently, sufficient thermonuclear energy must be released during each heating cycle to compensate for this degradation.

No claim that the above treatment is complete or applies to all possible types of system is made, but it does give some idea of the order of magnitude of the problems involved.

Systems which depart substantially from the electrically neutral Maxwellian gas assumed here have been carefully considered, but none looks promising. Some reasons for this are discussed by Thonemann (1956).

ACKNOWLEDGMENTS

This paper is an amplified version of the more important topics in A.E.R.E. report GP/R 1807. It owes much to discussions with several colleagues, in particular Mr. R. S. Pease, Dr. W. B. Thompson and Dr. P. C. Thonemann.

REFERENCES

POST, R. F., 1956, *Rev. Mod. Phys.*, **28**, 338.
SPITZER, L., 1956, *The Physics of Fully Ionized Gases* (New York : Interscience Publishers), p. 90.
THIRRING, H., 1955, *Nucleonics*, No. 11, p. 62.
THOMPSON, W. B., 1957, *Proc. Phys. Soc.* B, **70**, 1.
THONEMANN, P. C., 1956, *Nuclear Power*, **1**, 169.

Equilibrium Characteristics of a Pinched Gas Discharge Cooled by Bremsstrahlung Radiation

By R. S. PEASE
Atomic Energy Research Establishment, Harwell, Berks.

Communicated by D. W. Fry; *MS. received 5th November* 1956

Abstract. Standard formulae for the electrical resistance and for the radiating properties of a fully ionized gas have been combined with pinch-effect relations to obtain the stationary state radial distribution functions and current–voltage characteristics of a radiation-cooled discharge, using a simplified model in which a stationary, axially symmetrical discharge is generated in a long straight tube. The validity of the diffusion conditions implied in the initial equations, the transport properties of the plasma, and the cyclotron radiation arising from thermal motion in the self-magnetic field, are considered. The calculations are applied to the requirements of a power-producing thermonuclear reactor.

The chief results are that in a stationary state a pinched discharge will exist with radiation cooling, that the magnitudes of the voltages, tube diameters and initial gas pressure are within reasonable engineering limits, and that a maximum current of one to two million amperes will be encountered. This maximum is a characteristic current of an ionized gas and, slowly varying numerical factors apart, depends directly only on the fundamental constants $\boldsymbol{e}$, $\boldsymbol{m}$, $\boldsymbol{h}$ and $\boldsymbol{c}$.

While it is possible to present grounds for supposing these results to be independent of some of the assumptions, the work gives no guidance on the extent to which they will be invalidated by the instabilities of the current channel.

§ 1. INTRODUCTION

To obtain useful power from thermonuclear reactions of hydrogen isotopes, it is necessary (Lawson 1957) to achieve temperatures in the range 10^7–10^9 deg K, and to do so without dissipating amounts of heat which are large compared to the inescapable radiation losses. One promising method is that of the pinched gas discharge; and the purpose of the present work is to explore this possibility by calculating, in an elementary fashion, the characteristics of a hydrogen discharge in which dissipation of energy is supposed to take place entirely by radiation of bremsstrahlung.

As will be seen, many simplifying assumptions have been made, either to avoid mathematical complexity, or from lack of information on the relevant physical phenomena. Within the limitations imposed by the assumptions, the calculations suggest that discharges which are sufficiently pinched for wall losses to be negligible, and of sufficient temperature to produce thermonuclear reactions economically, can exist without the employment of prohibitively large electric fields or apparatus of astronomical dimensions. A further result is that there appears to be a maximum equilibrium current of about 10^6 amperes which it will not be possible to exceed in the anticipated physical conditions however large the applied electric field, without introducing non-stationary conditions or otherwise introducing physical phenomena not covered by the equations. It should be emphasized that the calculations apply to a stable discharge, and assume that a stationary state has been reached. The practical application of the optimistic results obtained is likely to depend essentially both on the suppression of instabilities and on the ability to get the discharge into the stationary radiation-cooled configuration.

Most published treatments of the pinch effect have dealt mainly with the balance of the self-magnetic force by pressure or inertial forces (Bennet 1934, Tonks 1937, 1939, Schlüter 1950, Thonemann and Cowhig 1951, Blackman 1951, Leontovitch and Osovets 1956). To define the discharge conditions for a steady state it is also necessary to consider how the energy fed into the gas is dissipated. Recently this has been done by Thonemann and by Ware (private communications) for the case of dissipation by thermal conduction to the walls of the discharge tube. However, if the discharge becomes sufficiently constricted, this conduction cannot take place, and dissipation by radiation will become important. The possibility of a current limitation due to the pinch effect has been discussed by Tonks (1937) and Alfvén (1950); their work would now appear to refer to the conditions at which a transition to the radiation-cooled configuration might occur. Work showing some features in common with that given here has been published by Budker (1956) for the case of a pinched beam of relativistic electrons passing through a low pressure gas, where the energy dissipation arises mainly from cyclotron radiation. The present work has been reported briefly (Pease 1956).

None of the published experimental studies (Ware 1951, Thonemann and Cowhig 1951, Artsimovitch *et al.* 1956) would appear to provide observational data to test the calculations given here.

§ 2. ASSUMPTIONS

The discharge is supposed to take place in a straight tube of radius a, containing $\pi a^2 \bar{n} = N$ hydrogen atoms per unit length, under the influence of a uniform applied axial electric field E. The tube is supposed to be sufficiently long for end effects to be neglected, and the calculations are only concerned with the plasma, where the electron density n and the temperature T are assumed equal to the ion density and temperature respectively. The walls specularly reflect any protons or electrons which impinge upon them, but absorb all the radiation emitted by the discharge which itself is optically 'thin' to the bulk of the ordinary bremsstrahlung. The hydrogen is completely ionized, and the current is carried by the electrons which drift with velocity w. Co-operative phenomena such as plasma oscillations and instabilities are assumed to be absent, and the discharge is assumed to be cylindrically symmetrical about the axis of the tube.

The essential physical process of concern in this paper is the acquisition by the electrons of momentum and energy from the electric field, and the dissipation of these by Coulomb collisions with the protons. The main effect of collisions between electrons and electrons, and between protons and protons is to redistribute the energy and momentum. These thermal conduction and viscous effects are treated here by calculating results for extreme values of the conductivity and viscosity, and by subsequent estimation of the transport processes.

Diffusion conditions with Maxwellian distribution of velocities are assumed. This assumption requires that at each point in the plasma

$$\lambda_\perp \frac{dX}{dr} \ll X \quad \text{and} \quad w \ll \left(\frac{kT}{m}\right)^{1/2}.$$

Here $\lambda_\perp$ is the mean radial separation of successive electron collisions; X is any of the variables n, w and T, m is the electron mass and k is Boltzmann's constant. The physical identification of $\lambda_\perp$ and the validity of these assumptions is discussed later. Unrationalized m.k.s. units are used.

§ 3. BASIC EQUATIONS

3.1. *The Dissipation of Energy*

Using the results of Gaunt (1929) on the Born approximation for electron–proton collisions, Cillie (1931) obtains for the power radiated per unit volume, $Bn^2T^{1/2}$; B is a constant whose numerical value is $1{\cdot}4\times10^{-40}$ m.k.s. units. In equilibrium, the power delivered electrically, EI, must be equal to that radiated, and so

$$EI = \int_0^a 2\pi r B n^2 T^{1/2}\,dr; \qquad \ldots\ldots(1)$$

I is the total current and r is the distance from the axis of the discharge tube.

3.2. *The Balance of Axial Momentum*

According to diffusion theory (e.g. Chapman and Cowling 1939) the momentum transferred per unit time to the protons by electrons drifting with a velocity w is $n^2e^2wT^{-3/2}G\log\Lambda$. Here G is a constant and Λ is essentially the ratio of the maximum impact parameter to the wavelength of the electrons. In the stationary state, the total axial momentum of the electrons is constant, and so:

$$EeN = \int_0^a 2\pi r n^2 e^2 w T^{-3/2} G \log\Lambda\,dr. \qquad \ldots\ldots(2)$$

The constant G is equal to $(4e^2/3\epsilon_0^2)(2\pi m/k^3)^{1/2}$ and has the value 129 m.k.s. units (Spitzer 1956, Eqn 5.39); ϵ_0 is the dielectric constant of free space. The term $\log\Lambda$ is $\ln[kT(3m\epsilon_0)^{1/2}/e\hbar\sqrt{(\pi n)}]$ in the present case of Born approximation and Coulomb cut-off at the Debye distance; it has a value of about 10 and is so insensitive to variation of n and T that it can be regarded as a constant. The use of these results for the case of current flow perpendicular to a magnetic field is justified in this case of the self-magnetic field (Alfvén 1950, Spitzer 1952) essentially because the corresponding Hall current cannot flow.

3.3. *The Pinch Effect*

The kinetic pressure gradient of the plasma is everywhere balanced by the magnetic force produced by the axial flow of current. Using i to denote the current enclosed by the cylinder of radius r, the magnetic field strength is $2i/r$, and there is also the relation:

$$\frac{di}{dr} = 2\pi r n e w. \qquad \ldots\ldots(3)$$

The kinetic pressure is equal to $2n\boldsymbol{k}T$, and consequently the balance of pressure is expressed by

$$\mu_0 \frac{i\,di}{dr} + 2\pi r^2 \frac{d}{dr}(n\boldsymbol{k}T) = 0, \qquad \ldots\ldots(4a)$$

where μ_0 is the permeability of free space. Equation (4*a*) can be integrated directly to give

$$\mu_0 i^2 + 4\pi r^2 n\boldsymbol{k}T = \int_0^r 8\pi r n \boldsymbol{k} T\,dr. \qquad \ldots\ldots(4b)$$

If the discharge is well pinched, the term $4\pi r^2 n\boldsymbol{k}T$ is negligible at the walls of the tube, and equation (4*b*) becomes

$$\mu_0 I^2 = 4N\boldsymbol{k}\overline{T} \qquad \ldots\ldots(4c)$$

where $\overline{T}$ is the mean temperature obtained by averaging over all charged particles per unit length of the discharge.

§ 4. SOLUTIONS OF THE EQUATIONS

The solutions required are the total current I, and the quantities n, T, w and i as functions of r and of the controllable quantities E, a and $\overline{n}$. As indicated above, solutions cannot be obtained without further equations introducing transport processes, arising from variations of T and w with r. In the present work, the solutions are obtained for several cases where the transport processes introduce no mathematical complexity.

Case (*i*): *Zero Viscosity and Zero Thermal Conductivity.*

Since there is no transport of momentum or energy from one part of the discharge to another, equations (1) and (2) reduce to

$$E\,di = 2\pi r n^2 T^{1/2} B\,dr \qquad \ldots\ldots(1a)$$

and

$$E = new T^{-3/2} G \log \Lambda. \qquad \ldots\ldots(2a)$$

Combining these with (3) and (4*b*) yields

$$\frac{di}{dr} = 2i\,\frac{P-i}{Pr} \quad \text{where} \quad P = \frac{4\boldsymbol{k}}{\mu_0}\left(\frac{G\log\Lambda}{B}\right)^{1/2}.$$

It follows that i and hence I, cannot exceed P. Introducing a parameter A_1 by means of the equation $I = A_1 P/(1+A_1)$ the following distribution equations are obtained:

$$i = P\left(1 + \frac{a^2}{A_1 r^2}\right)^{-1} \qquad \ldots\ldots(5a)$$

$$n = \tfrac{1}{3}\overline{n}\left(1 + \frac{A_1 r^2}{a^2}\right)^{-2/3} A_1[(1+A_1)^{1/3} - 1]^{-1} \qquad \ldots\ldots(5b)$$

$$T = \frac{\mu_0 P^2}{4\boldsymbol{k}N}\left(1 + \frac{A_1 r^2}{a^2}\right)^{-4/3} 3[(1+A_1)^{1/3} - 1] \qquad \ldots\ldots(5c)$$

$$w = \frac{P}{Ne}\left(1 + \frac{A_1 r^2}{a^2}\right)^{-1/3} 3[(1+A_1)^{1/3} - 1]. \qquad \ldots\ldots(5d)$$

The form of these equations shows that n, T and w all have maximum values at $r=0$, and fall off towards the walls of the discharge tube. The parameter A_1 is, roughly speaking, the ratio of the tube cross section to the current channel cross section, indeed the magnetic field is a maximum at $r = a/\sqrt{A_1}$. The relation between the field strength and this 'pinch parameter' A_1 is

$$E = A_1 E_0 [108\{(1+A)^{1/3} - 1\}^3]^{-1/2}, \qquad \ldots\ldots(6)$$

where

$$E_0 = (B/\pi a^2)(\mu_0 N^3/k)^{1/2}.$$

Case (ii): Zero Viscosity and Infinite Thermal Conductivity.

This is the case of constant current density within a well-defined channel of radius b; and it is only necessary to combine equation (1) with published analyses (e.g. Schlüter 1950). Two solutions must be considered: one in which $b < a$, and one in which $a < b$. For the first solution, putting $a^2 = A_2 b^2$, and $n = 0$ when $r \geqslant b$,

$$n = 2\bar{n} A_2 (1 - A_2 r^2/a^2) \qquad \ldots\ldots(7a)$$

and

$$w = \tfrac{1}{2}\bar{w}(1 - A_2 r^2/a^2)^{-1}. \qquad \ldots\ldots(7b)$$

Equations (7) hold only inside the current channel defined by $r = a/\sqrt{A_2}$. Outside this radius there is a complete vacuum. Consequently, equation (4*c*) holds rigorously. Combining equations (7), (1) and (4*c*), the current is found to be

$$I = \frac{k}{\mu_0}\left[\frac{12 G \log \Lambda}{B}\right]^{1/2} = \sqrt{\tfrac{3}{4}}\,P \qquad \ldots\ldots(8a)$$

for all values of E, n, a provided $A_2 > 1$. Eliminating I and T from the same equations yields

$$E = \tfrac{2}{3} A_2 E_0. \qquad \ldots\ldots(9a)$$

For the solution when $b > a$, the same parabolic distributions (7) hold, but of course n is not zero at the walls. The expressions for the current and voltage in terms of the parameter A_2 are

$$I = (\tfrac{3}{4})^{1/2} P A_2 (3 - 3A_2 + A_2^2)^{-1/2} \qquad \ldots\ldots(8b)$$

$$E = \tfrac{2}{3} E_0 (2 - A_2)^{3/2} A_2^{-1/2} (3 - 3A_2 + A_2^2)^{-1}. \qquad \ldots\ldots(9b)$$

Case (iii): Infinite Viscosity and Infinite Thermal Conductivity.

In this case both w and T are independent of radius. From the analysis of Bennet (1934), n is given by

$$n = \bar{n}(1 + A_3)(1 + A_3 r^2/a^2)^{-2}, \qquad \ldots\ldots(10)$$

where A_3 is a constant to be determined. The current–voltage characteristics are obtained in terms of A_3 from equations (1)–(4):

$$I = P\frac{A_3}{1 + A_3} \qquad \ldots\ldots(11)$$

$$E = \tfrac{1}{2} E_0 \left[1 + \frac{A_3^2}{3(1 + A_3)}\right][1 + A_3^{-1}]^{1/2}. \qquad \ldots\ldots(12)$$

Case (iv): Zero Thermal Conductivity and Infinite Viscosity.

The starting equations are (3), (4*a*), (1*a*), and equation (2) simplified by w being independent of r. Equations (3) and (1*a*) yield $nT^{1/2} = Eew/B$, which is thus also independent of r. The analysis is carried out in terms of the solution $\tau(x)$ of the equation:

$$\tau^{1/2}\frac{d}{dx}\left(x\frac{d\tau}{dx}\right) + x = 0 \qquad \ldots\ldots(13)$$

for the boundary conditions $\tau=1$, $d\tau/dx=0$, when $x=0$. The temperature is given by $T=T_0\tau(rA_4^{1/2}/a)$, where T_0 is the temperature at $r=0$, and A_4 is a pinch parameter analogous to those of cases (i)–(iii). The number and current density, both of which increase towards the walls of the tube, are obtained from the relation $nT^{1/2}=E\epsilon w/B$. The total current and the electric field are:

$$I=\frac{1}{2}\frac{\boldsymbol{k}}{\mu_0}\left(\frac{G\log\Lambda}{B}\right)^{1/2}[F_1(A_4)\,F_2(A_4)]^{1/2} \qquad \ldots\ldots(14)$$

and

$$E=E_0A_4/[F_1(A_4)]^{3/2}\sqrt{2} \qquad \ldots\ldots(15)$$

where $F_1(A_4)\equiv\int_0^{\sqrt{A_4}}\tau^{-1/2}x\,dx$ and $F_2(A_4)\equiv\int_0^{\sqrt{A_4}}x\tau^{-5/2}\,dx.$

The form of equations (14) and (15) is similar to the equivalent equations for cases (i)–(iii). However, in this case there is no finite solution for $A_4\geqslant 3{\cdot}35$. When $A_4=3{\cdot}35$, $F_2(A_4)$, and hence the equilibrium current, is infinite. Numerical values of τ, F_1 and F_2 have been kindly worked out by Mr. E. B. Fossey, and are presented in the table.

Solutions of Equation (13)

x	$\tau(x)$	$F_1(x^2)$	$F_2(x^2)$
0	1·000	0	0
0·4	0·960	0·081	0·084
0·8	0·837	0·334	0·400
1·2	0·621	0·803	1·31
1·6	0·288	1·647	6·17
1·8	0·051	2·527	78·5
1·83	(0·007)	2·865	(∞)

§ 5. Features of the Four Cases and Numerical Values

In all four cases a parameter A_i has been introduced, which is the ratio of the area of the cross section of the discharge tube to a characteristic area of the discharge. In each case, the total current depends only on the value of A_i and on the characteristic current P of these discharges. This current has the value $1{\cdot}7\times10^6$ amperes, where $\log\Lambda$ has been taken to be 10. In terms of fundamental constants:

$$P=2[3^{1/2}\log\Lambda]^{1/2}\boldsymbol{m}\boldsymbol{e}^{-2}\hbar^{1/2}\mathbf{c}^{7/2}\ \text{(e.s.u.)}$$

which, like all currents depending only on $\boldsymbol{e}$, $\boldsymbol{m}$, $\hbar$ and $\mathbf{c}$, can be written $(\boldsymbol{m}\mathbf{c}^3/\boldsymbol{e})f(\alpha)$, where $f(\alpha)$ is any function of the fine structure constant α, which in this case is $\alpha^{-1/2}$.

The relations between I and A are illustrated in figure 1. As can be seen the results for cases (i)–(iii) are very similar; in particular a current, which is exactly P in (i) and (iii) and is only slightly less for case (ii), is reached which is not exceeded however high A is made, apart from very slow variations arising from the term $\log\Lambda$. In case (iv), no finite solution has been found for large values of A; however, as is shown below, this case is a very artificial one from various points of view. These current characteristics are altered only by a factor $(Z+1)/2Z$ if ions of charge Ze replace the protons assumed here. The experimental variables E, a and $\bar{n}$ only affect the current in so far as they are related to A through the characteristic field strength $E_0=(B/\pi a^2)(N^3\mu_0/\boldsymbol{k})^{1/2}$. The relations between E, E_0 and A are illustrated in figure 2, and are similar in all four cases. The result for case (i) is exceptional because the number for electrons carrying the bulk of the current decreases as $A^{1/3}$ (cf. equations (5*a*) and (5*b*)). By associating N with this number rather than with the total number of electrons in the tube, the relation between E and A for case (i) can be brought into line with the other cases.

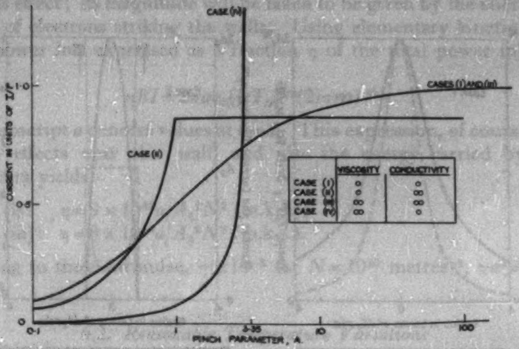


Figure 1. Current as a function of pinch parameter A. The characteristic current P has a value of about $1 \cdot 7 \times 10^6$ ampere.

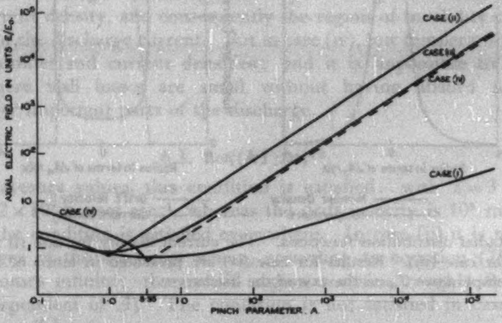


Figure 2. The relation between pinch parameter A and axial electric field. The characteristic field strength E_0 is equal to $(B/\pi a^2)(\mu_0 N^3/k)^{1/2}$. The curve for case (iv) is dashed for $A > 3 \cdot 35$, because the equilibrium current and temperature are infinite.

To fix ideas, both here and in what follows, the values of the experimental variables are supposed to be: $\pi a^2 = 1$ metre^2; $N = 10^{20}$ hydrogen atoms per metre length. These yield $E_0 = 0 \cdot 012$ volt per metre, so that, in principle, high values of A are easily obtainable. For $N = 10^{20}$, equation (4*c*) yields $T = 3 \times 10^7$ deg K for a current of P.

Diagrams of the radial distributions are shown at figure 3. The individual quantities, such as n, w and current density new, are not always a maximum on the axis; nor do they necessarily become zero at the walls for sufficiently high A (e.g. n in case (i)). But in all cases, the kinetic energy density $3nkT$ is a maximum on the tube axis, and can be reduced to an indefinitely low value at the walls of the tube by applying a suitable electric field. The gas pressure is then entirely supported by the self-magnetic field, and in this sense the discharges are pinched in each case. When the discharge is well-pinched its characteristics must be independent of the walls, and the radius b of the discharge is related to the electric field E in terms of fundamental constants by

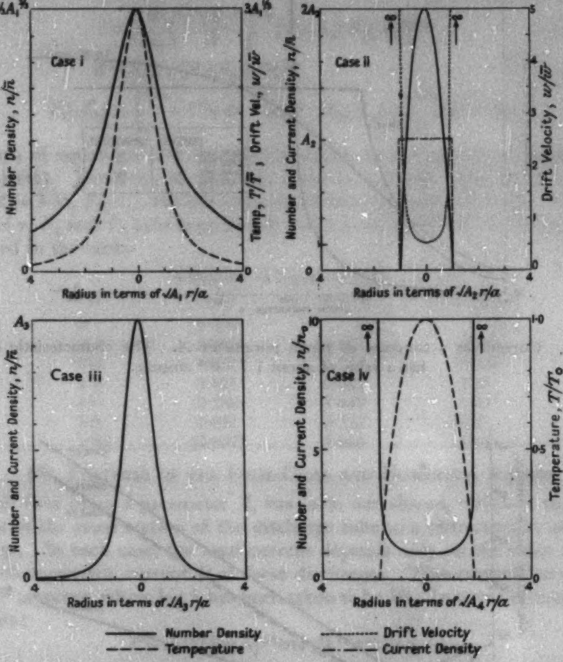


Figure 3. Radial distribution functions. The current density for case (i) is the same as that for case (iii). Results for case (iv) are presented in terms of the density n_0 and temperature T_0 on the axis of the discharge.

$$E=\left(\frac{\mathbf{e}}{b^2}\right)\left(\frac{\mathbf{e}^2}{hc}\right)\left(\frac{r_0}{l}\right)^{3/2}\left(\frac{2^{13}\pi}{3^5}\right)^{1/2},$$

where the numerical factors are appropriate to case (ii), $\mathbf{e}$ is in e.s.u., r_0 is the classical radius of the electron and $l=N^{-1}$.

§ 6. EXAMINATION OF ASSUMPTIONS

6.1. *Wall Losses*

In practice, dissipation of momentum and energy at the walls will be negligible only if the value of nT is sufficiently small at the walls. Energy loss is much the more serious effect; its magnitude will be taken to be given by the energy carried by the flux of electrons striking the walls. Using elementary kinetic theory of gases, the power loss expressed as a fraction η of the total power input EI is given by

$$\eta EI = 2\pi a n_a (\boldsymbol{k}T_a)^{3/2}(2/\pi \boldsymbol{m})^{1/2}$$

where the subscript a denotes values at $r = a$. This expression, of course, neglects any sheath effects near the wall, and also the energy carried by protons. Simple algebra yields:

case (i): $\eta = 9 \times 10^{49} a / A_1^2 N^2$ (m.k.s.);
case (iii): $\eta = 3 \times 10^{49} a / A_3^2 N^2$ (m.k.s.).

According to these formulae, $\eta \simeq 10^{-3}$ for $N = 10^{20}$ metres^{-1}, $\pi a^2 = 1$ metre^2, and $A = 10^6$.

6.2. *Reasonable Temperature Variations*

If the temperature is less than 5×10^5 deg K, equation (1) breaks down; if it exceeds 5×10^9 deg K, the electron velocities are relativistic. Since the current is fixed, this restricts N through equation (4*c*) to the range 10^{18}–10^{22} per metre. The restriction is more severe for cases (i) and (iv) where the temperature varies throughout the discharge. In case (i), the very low temperature region coincides with low current density, and consequently the region of invalidity carries only a small part of the discharge current. But in case (iv), low temperatures coincide with high number and current densities; and it is impossible to conceive a situation where wall losses are small without having absurd temperature variations over important parts of the discharge.

6.3. $w \ll (\boldsymbol{k}T/\boldsymbol{m})^{1/2}$

For the average values, this condition is satisfied: with $T = 3 \times 10^7$ deg K, $(\boldsymbol{k}T/\boldsymbol{m})^{1/2} = 2 \times 10^7$ metres sec^{-1}, whereas the drift velocity is 10^5 metres sec^{-1}. In case (i), the condition is satisfied everywhere. In case (ii) it is not satisfied near the edges of the discharge where, according to equation (7*b*) the drift velocity becomes infinite; the invalid region contains roughly 0·1% of the current (independent of A). The condition is not satisfied in case (iv), near the edge where T becomes zero.

6.4. $\lambda_\perp dX/dr \ll X$

The mean free path λ for momentum transfer in electron–electron and electron–ion collisions is readily calculated, using the work of Chapman and Cowling (1939) and others, to be about 1 cm for a temperature 10^8 deg K and $n = 10^{26}$ m^{-3}. However, the radius of gyration R_L in a magnetic field H, which is $(2\boldsymbol{m}\boldsymbol{k}T)^{1/2}/eH\mu_0$, is about 10^{-6} metre for 10^8 deg K and $H = 10^6$ oersteds. This latter distance is small compared both with λ and with the radius of the discharge. Consequently, an electron must make many collisions to move radially a distance large compared with R_L. Therefore the quantity $\lambda_\perp$ of §2 has been identified with R_L; consequently for case (i) it is easily found that

$$\frac{R_L}{T}\frac{dT}{dr} = \frac{R_L}{w}\frac{dw}{dr} = \frac{4(2\boldsymbol{mkT})^{1/2}}{3P\mu_0\boldsymbol{e}},$$

for all values of r. This has the value 2×10^{-3} for $T=10^8$. Similarly, it is easily shown that the variations of n over the distance R_L are small; and it can likewise be shown that the conditions are satisfied in the other cases, with $A\simeq10^6$, except near infinities or zeros at the edges. For the bulk of the discharge this result, as in case (i), is independent of the value of A.

§ 7. MAGNITUDE OF THE TRANSPORT PROCESSES

For the radial directions, the transport coefficients of viscosity and thermal conductivity of an electron gas are reduced from their values in zero magnetic field by a factor of about $[1+(\lambda/R_L)^2]^{-1}$ (Chapman and Cowling 1939, pp. 179, 335–8). For the conditions mentioned above, this factor is much less than unity provided the magnetic field is greater than about 100 oersteds. Evidently then, the transport equations are governed by R_L over all but a tiny region at the centre of the discharge, and the approximate form $(R_L/\lambda)^2$ of the above factor is used below.

7.1. *Viscosity*

The electric force per electron $\boldsymbol{Ee}$ has been compared with the viscous force per electron,

$$\frac{1}{nr}\frac{d}{dr}\left(r\mu\frac{dw}{dr}\right)$$

where μ is the coefficient of viscosity and dw/dr is calculated assuming zero viscosity. Using the formulae for a simple electron gas,

$$\mu=\left(\frac{1}{10}\right)\left(\frac{ren}{i\mu_0\epsilon_0}\right)^2\left(\frac{\pi\boldsymbol{m}^3}{\boldsymbol{k}T}\right)^{1/2}\log\Lambda.$$

Substitutions in the results of §4 yield:

$$\frac{1}{\boldsymbol{E}\boldsymbol{e}nr}\frac{d}{dr}\left(\frac{\mu r\,dw}{dr}\right)=\left(\frac{3\sqrt{2}}{20}\right)\left(\frac{\boldsymbol{m}}{\mu_0\boldsymbol{e}^2N}\right)A^{1/3}(1+Ar^2/a^2)^{-4/3}, \text{ for case (i);}$$

$$=0 \text{ for case (ii).}$$

For case (i), over the bulk of the discharge where $Ar^2/a^2\simeq1$, the electric force is about 10^4 times the viscous force for $A=10^6$. In case (ii), the factor $(d/dr)(r\mu\,dw/dr)$ is exactly zero everywhere; an order of magnitude calculation with a distorted solution leads to the same conclusion as for case (i).

7.2. *Thermal Conductivity*

The heat flow from an annular element of thickness dr and unit length is $2\pi\,dr(d/dr)(r\kappa\,dT/dr)$, where κ the thermal conductivity is given by $(3\boldsymbol{k}/4\boldsymbol{m})5\mu$. The heat radiated is $2\pi r\,dr Bn^2T^{1/2}$. The ratio of these two quantities has been calculated for case (i), and found to be a pure number close to unity. Consequently, thermal conduction by the electrons may well be important. Moreover, the conduction of heat by the protons is likely to be much larger because of the much larger radius of gyration.

It thus appears that case (ii), zero viscosity and infinite thermal conductivity, is likely to be the best simple approximation to use. In the circumstances envisaged, the case of large viscous transfer and negligible thermal conduction is highly implausible.

§ 8. CYCLOTRON RADIATION

The energy W_M radiated by the electrons during their gyrations in the magnetic field is (Alfvén 1950):

$$W_M = \int_0^a 2\pi r n \mathbf{k} T H^2 M \, dr, \qquad \ldots\ldots(16)$$

where M is equal to $4\mathbf{e}^4\mu_0^3/3\mathbf{m}^3\mathbf{c}$, and has the value $3{\cdot}9 \times 10^{-15}$ m.k.s. units. Applying this result to case (ii), it is found that W_M is about a tenth of the total power EI. This proportion depends only on N, and when $N = 2{\cdot}5 \times 10^{19}$ metre^{-1} is equal to unity. However, Spitzer (private communication) has suggested that the discharge will not be optically thin to magnetic bremsstrahlung. A calculation of discharge characteristics based on (16) instead of (1), leads to the conclusion that the ratio of W_M to the black body radiation in the frequency range of zero to the maximum gyro-frequency, would be about $\mathbf{c}/\bar{w}$. Consequently (16) could only represent power leaving the plasma when $\bar{w}$, the mean drift velocity, approaches the velocity of light, as is so in the case treated by Budker (1956).

Thus the chief role of cyclotron radiation in the present case is to provide an additional mechanism for the transfer of energy from one part of the discharge to another, and hence is an additional reason for accepting case (ii) as the best simple representation of the transport processes. For the numerical values used above the maximum cyclotron frequency, which occurs at the edge of the discharge, is 2×10^{13} sec^{-1}; the maximum characteristic plasma frequency is about 10^{14} sec^{-1}, in the centre of the current channel.

§ 9. DISCUSSION

The results of the calculations may be summarized as follows:

(*a*) The radiation cooling in a stationary state permits a pinched discharge to exist. For a given number of particles per unit length in the current channel, the discharge cross section is inversely proportional to the applied field.

(*b*) The current is only a function of the pinch parameter A and of a characteristic current $(4\mathbf{k}/\mu_0)[G(\log\Lambda)/B]^{1/2}$ which has the value $1{\cdot}7 \times 10^6$ amperes. For $A \gg 1$, the current is virtually independent of A. The reduction of the discharge cross section as E is increased reduces the conductance of the channel so as to leave the current unchanged. The additional power is radiated as a consequence of the increase in the mean square number density.

(*c*) The estimates of the transport processes suggest that over the bulk of the discharge, viscous forces tending to keep the electron drift velocity constant are small, while thermal conduction tending to keep the temperature constant is appreciable. Consequently, of the four cases worked out, case (ii) ($\mu = 0$, $\kappa = \infty$) is likely to be nearest the actual situation.

(*d*) The assumption of diffusion conditions implicit in the starting equations are valid for case (ii) except in a limited region at the edge of the current channel.

The chief restrictions on validity are $10^{18} < N < 10^{22}$ particles per metre, and $E \gg E_0$ so that cooling by conduction to the walls is negligible.

(*e*) The rate at which cyclotron radiation is generated becomes large for $N < 2 \times 10^{19}$ metres^{-1}, but most of the power cannot escape and is reabsorbed in the plasma.

These conclusions must be regarded with some reserve on account of factors omitted from the equations. Examples are: cylindrical symmetry has been assumed; cooperative phenomena have been neglected; the ion temperature is assumed equal to the electron temperature; cross-terms (Ware 1954) in the transport processes have been neglected. However, the conclusion that the discharge will be pinched is not inherently implausible. This conclusion is simply that the discharge does not have to become heated to an 'unpinchable' temperature in order to dissipate the energy fed in. This arises essentially from the binary collision rate factor of n^2 in Cillie's radiation equation, which ensures that the power dissipated increases rapidly with the degree of pinch. This conclusion is therefore unlikely to depend on the details of the assumptions in the calculations.

The conclusion that the current will be limited to a more or less fixed value of $1{\cdot}7 \times 10^6$ amperes is more unexpected; and clearly from the slight difference between case (ii) and cases (i) and (iii), and from the results obtained in the implausible conditions of case (iv), this conclusion does depend to some extent on the details of the conditions. The generality of the result can be indicated as follows. An electron which drifts an axial distance D absorbs momentum EeD/w and energy EeD from the axial electric field. In a steady state, this is dissipated by collisions, in each of which on the average a fraction f_1 of the energy $\boldsymbol{k}T$ and a fraction f_2 of drift momentum mw is lost; it follows that, whatever E, $mw^2/\boldsymbol{k}T = f_1/f_2$ (see for instance Thonemann 1956); but if the discharge is well-pinched, the relation (4*c*) holds, which serves to define T in terms of I. Consequently, if f_1/f_2 is variable only in so far as T is variable, the equations define unique values of I, independent of E, which the pinched discharge must adopt to be in a stationary state. This result does not depend upon peculiarities of Coulomb interaction; this produces only the particular dependence on T which also eliminates from the equations the number of ions per unit length. The pinch parameter affects the current only in so far as it affects the accuracy of equation (4*c*) as an approximation to (4*b*), that is in so far as the pinch is not perfect. This simple deduction is based on the use of constant w and T; clearly when w and T vary with radius, average values must be taken, and these depend on details of the discharge conditions.

The most obvious and serious obstacle to achieving discharge conditions in which these calculations are applicable lies in the instabilities (Kruskal and Schwarzschild 1954, Tayler 1957) known to exist in current-carrying plasmas. These may well preclude the existence of a true steady state. It seems possible that a steady or quasi-steady state might be obtained by applying an axial magnetic field, as discussed for instance by Artsimovitch (1956). In a true steady state where the axial field is uniform, the anisotropic conductivity of the plasma in a magnetic field (Spitzer 1952) will have to be allowed for. In a quasi-steady state, the axial field outside the discharge might be larger than that inside, in which case the current channel will be compressed, and hence the current would be less than that calculated here. Alternatively, the current channel could be expanded by a trapped axial field, and the current would be greater.

ACKNOWLEDGMENTS

This work owes much to the patient instruction of Dr. P. C. Thonemann and other colleagues. I am particularly indebted to Dr. J. E. Allen and Dr. R. J. Tayler for checking much of the algebra and for constructive criticism of the manuscript. I am indebted to Mr. R. T. P. Whipple and Mr. E. B. Fossey for the solution of equation (13).

REFERENCES

Alfvén, H., 1950, *Cosmical Electrodynamics* (Oxford : University Press)
Artsimovitch, L., 1956, I.A.U. Symposium, Stockholm.
Artsimovitch, L., Andrianov, A., Bazilevskaya, O., Prokhorov, Y., and Fillipov, N., 1956, *Atomnaya Energiya*, **3**, 76.
Bennet, W. H., 1934, *Phys. Rev.*, **45**, 890.
Blackman, M., 1951, *Proc. Phys. Soc.* B, **64**, 1039.
Budker, G. J., 1956, C.E.R.N. Symposium on High Energy Accelerators and Pion Physics (Geneva).
Chapman, S., and Cowling, T. G., 1939, *The Mathematical Theory of Non-Uniform Gases* (Cambridge : University Press).
Cillie, C., 1931, *Mon. Nat. R. Astr. Soc.*, **92**, 820.
Gaunt, J. A., 1929, *Phil. Trans. Roy. Soc.* A, **229**, 200.
Kruskal, M., and Schwarzschild, M., 1954, *Proc. Roy. Soc.* A, **223**, 348.
Lawson, J. D., 1957, *Proc. Phys. Soc.* B, **70**, 6.
Leontovitch, M., and Osovets, S., 1956, *Atomnaya Energiya*, **3**, 81.
Pease, R. S., 1956, I.A.U. Symposium, Stockholm.
Schlüter, A., 1950, *Z. Naturforsch.*, **5a**, 72.
Spitzer, L., 1952, *Astrophys. J.*, **116**, 299; 1956, *Physics of Fully Ionized Gases* (New York : Interscience).
Tayler, R. J., 1957, *Proc. Phys. Soc.* B, **70**, 31.
Thonemann, P. C., 1956, *Appl. Sci. Res.* B, **5**, 237.
Thonemann, P. C., and Cowhig, W. T., 1951, *Proc. Phys. Soc.* B, **64**, 345.
Tonks, L., 1937, *Trans. Electrochem. Soc.*, **72**, 167; 1939, *Phys. Rev.*, **56**, 360.
Ware, A. A., 1951, *Phil. Trans. Roy. Soc.* A, **243**, 863; 1954, *Proc. Phys. Soc.* B, **67**, 869.

An Elementary Theory of the Transient Pinched Discharge

By J. E. ALLEN

Atomic Energy Research Establishment, Harwell, Berks.

Communicated by D. W. Fry ; *MS. received* 1*st November* 1956

Abstract. A simple theory of the collapse of a high current discharge plasma is described. The theory applies to discharges of the type studied by Cousins and Ware, in which shock waves are generated by the self-magnetic 'pinch' forces.

---

§ 1. Introduction

Experiments have shown that if a high current is rapidly established in a low pressure gas discharge, the plasma is accelerated towards the discharge axis by the self-magnetic 'pinch' force (Cousins and Ware 1951, Artsimovich *et al.* 1956, and unpublished work carried out in this laboratory). This magnetic pinch force results from the interaction of the discharge current with its own magnetic field. The pinch effect is now well known, but the only published treatment referring to the development of the constricted channel is the very recent one by Leontovich and Osovets (1956). The theory described in the present paper differs from that of Leontovich and Osovets and refers to the case where shock waves are set up in the gas.

It is assumed, for simplicity, that the discharge consists of a large slab of completely ionized gas (as shown in figure 1), in which the following conditions exist: (*a*) the electrical conductivity is infinite, (*b*) the current rises instantaneously from zero to a finite value and then remains constant, and (*c*) the thickness of the shock wave is small compared with the dimensions of the discharge chamber.

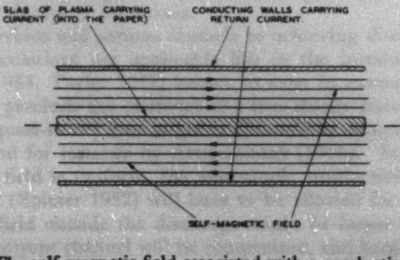

SLAB OF PLASMA CARRYING CURRENT (INTO THE PAPER)
CONDUCTING WALLS CARRYING RETURN CURRENT
SELF-MAGNETIC FIELD

Figure 1. The self-magnetic field associated with a conducting slab of gas.

§ 2. THEORY OF THE TRANSIENT PINCHED DISCHARGE

2.1. *Mechanism of Energy Transfer*

It follows from the first condition that only the energy delivered to the gas via electromechanical forces is considered in this theory; the energy dissipation associated with Joule heating is neglected. Another result which follows directly from the first condition is that the current flows only on the surface of the discharge, due to the skin effect, and the self-magnetic field does not penetrate the plasma. The 'pinch' force acting per unit area of the current sheet due to the interaction with its own magnetic field is $B^2/8\pi\mu$; this force is constant if the current remains constant, because $H=4\pi(I_S/c)$, where I_S is the sheet current (in e.s.u. per cm length). It is assumed, for simplicity, that the walls of the discharge chamber constitute the return current path (see figure 1). Hence the discharge inductance is zero, initially, so that conditions (*a*) and (*b*) are self-consistent.

Consider a time interval dt in which the current sheet moves a distance dx. The mechanical work done on the gas during this time is $(B^2/8\pi\mu)\,dx$ and is equal to the increment of energy stored in the magnetic field. The induced electric field ('back e.m.f.') resulting from the motion of the conducting sheet is $(B/c)\,dx/dt$ and the energy delivered by the electrical supply during the interval dt is $B(I_S/c)\,dx$, which is equal to $(B^2/4\pi\mu)\,dx$. Thus one half of the supplied energy is stored in the magnetic field and the other half is delivered to the gas via the electromechanical force. This result applies to all constant current systems.

2.2. *Inward-moving Shock Wave*

In the idealized case under consideration, a constant force drives the current sheet towards the centre of the discharge. A shock wave precedes the current sheet in the same way that a shock wave travels ahead of a piston which is plunged into a tube. The resulting sequence of events is illustrated in detail in the distance–time diagram (figure 2).

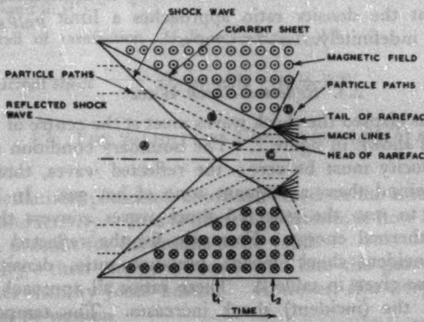


Figure 2. Distance–time diagram illustrating the sequence of events.

The inward-moving shock wave does two things: it sets the gas in motion, and it increases the temperature of the gas. Thus the energy which is delivered to the gas by the electromechanical force is partly associated with directed motion and partly with random (thermal) motion; in the limiting case of a strong shock, in a monatomic gas, the energy is equally divided between these. The gas is accelerated to the speed of the current sheet, and its pressure is increased from its initial value p_A to a value equal to the force per cm^2 acting on the current sheet, i.e. $p_B = B^2/8\pi\mu$. No net force is experienced by the piston after $t=0$ and it advances at constant velocity until it is hit by a reflected shock wave at $t=t_2$ (see figure 2). Formulae for the Mach number of the shock wave, the shock velocity, the density ratio and the temperature ratio are given in table 1 for the case of a monatomic gas. These are standard results of conventional shock wave theory (Courant and Friedrichs 1948).

Table 1. Inward-moving Shock Wave

		Limit for strong shocks
Mach number of the incoming flow relative to the shock front	$M^2 = \frac{1}{5}[4(p_B/p_A)+1]$	$M^2 \rightarrow \frac{4}{5}(p_B/p_A)$
Shock velocity in terms of velocity of current sheet (U_C)	$U = 2U_C/3 + [c_A{}^2 + (2U_C/3)^2]^{1/2}$	$U \rightarrow \frac{4}{3}U_C$
Density ratio	$\frac{\rho_B}{\rho_A} = \frac{4(p_B/p_A)+1}{(p_B/p_A)+4}$	$\frac{\rho_B}{\rho_A} \rightarrow 4$
Temperature ratio	$\frac{T_B}{T_A} = \frac{(p_B/p_A)[(p_B/p_A)+4]}{4(p_B/p_A)+1}$	$\frac{T_B}{T_A} \rightarrow \frac{5}{16} M^2$

As the electrical conductivity is infinite, the electromagnetic field is excluded from the discharge plasma. For this reason the generation and reflection of shock waves can be described in terms of ordinary hydrodynamics. It is of interest to note that the density ratio approaches a limit $\rho_B/\rho_A = 4$ as the current is increased indefinitely.

2.3. *Reflected Shock Wave*

After a time t_1 has elapsed the shock waves meet at the centre of the chamber and are reflected, as shown in figure 2. The boundary condition at the centre is that the flow velocity must be zero; the reflected waves, therefore, move outwards leaving behind them a stagnant zone of hot gas. In bringing the inward-moving gas to rest, the reflected shock waves convert the energy of mass motion into thermal energy. Formulae for the reflected shock speed (in terms of the incident shock speed), pressure ratio, density ratio and temperature ratio are given in table 2. These ratios all approach finite limits as the strength of the (incident) shock increases. The temperature ratio $T_C/T_B \rightarrow 12/5$ and the density ratio $\rho_C/\rho_B \rightarrow 5/2$, so that $\rho_C/\rho_A \rightarrow 10$. Again, the formulae are readily obtained using conventional shock wave theory.

2.4. *Expansion*

Consider the instant t_2 at which the reflected wave meets the inward-moving current sheet, as shown in figure 2. The gas is stationary and at a high temperature and pressure and, due to the increase in pressure associated with the reflected wave the gas pressure is greater than the magnetic force per cm^2 acting on the current sheet. As a result the latter is blown outwards and a rarefaction wave travels into the hot stagnant gas. The gas expands to a pressure $p_D = p_B = B^2/8\pi\mu$. Formulae for the pressure ratio, density ratio, temperature ratio, are given in table 3 together with expressions for the velocities of the head and tail of the rarefaction wave and the current sheet. These are based on an adiabatic expansion of the gas.

Table 2. Reflected Shock Wave

		Limit for strong shocks
$\dfrac{\text{Reflected shock speed}}{\text{Incident shock speed}}$	$\dfrac{U'}{U} = \dfrac{2+3(p_A/p_B)}{4+(p_A/p_B)}$	$\dfrac{U'}{U} \rightarrow \dfrac{1}{2}$
Pressure ratio	$\dfrac{p_C}{p_B} = \dfrac{6(p_B/p_A)-1}{(p_B/p_A)+4}$	$\dfrac{p_C}{p_B} \rightarrow 6$
Density ratio	$\dfrac{\rho_C}{\rho_B} = \dfrac{4(p_C/p_B)+1}{(p_C/p_B)+4}$	$\dfrac{\rho_C}{\rho_B} \rightarrow \dfrac{5}{2}, \ \dfrac{\rho_C}{\rho_A} \rightarrow 10$
Temperature ratio	$\dfrac{T_C}{T_B} = \dfrac{(p_C/p_B)[(p_C/p_B)+4]}{4(p_C/p_B)+1}$	$\dfrac{T_C}{T_B} \rightarrow \dfrac{12}{5}$

Table 3. Expansion

		Limit for strong shocks
Pressure ratio	$p_D/p_C = p_B/p_C$	$p_D/p_C \rightarrow 0{\cdot}167$
Density ratio	$\rho_D/\rho_C = (p_D/p_C)^{3/5}$	$\rho_D/\rho_C \rightarrow 0{\cdot}341$
Temperature ratio	$T_D/T_C = (p_D/p_C)^{2/5}$	$T_D/T_C \rightarrow 0{\cdot}488$
$\dfrac{\text{Velocity of head of rarefaction wave}}{\text{Velocity of sound in (C)}}$	$U_H/c_C = 1$	$U_H/c_C = 1$
Velocity of tail of rarefaction wave	$U_T/c_C = 3-4(p_D/p_C)^{1/5}$	$U_T/c_C \rightarrow 0{\cdot}205$
Velocity of current sheet	$U_C'/c_C = 3[1-(p_D/p_C)^{1/5}]$	$U_C'/c_C \rightarrow 0{\cdot}904$

§ 3. The Validity of Conditions (*a*), (*b*) and (*c*)

Condition (*a*).

The calculations presented above are based on the assumption of an infinite electrical conductivity. It is of interest to determine how high the conductivity must be in practice for the results to be valid.

If $Bv/c \gg j/\sigma$ then energy is transferred into the gas via electromechanical forces, rather than by Joule heating. However, $H/x \sim 4\pi j/c$, so that the above criterion can be written

$$v \gg \frac{c^2}{4\pi\sigma\mu x}. \qquad \ldots\ldots(1)$$

Now the equation describing the diffusion of a magnetic field through a conductor (the skin effect phenomenon) is

$$\nabla^2\mathbf{H} = \frac{4\pi\sigma\mu}{c^2}\frac{\partial\mathbf{H}}{\partial t}$$

and for an order of magnitude calculation we can put $\nabla^2\mathbf{H}\sim H/x^2$ and $\partial\mathbf{H}/\partial t\sim H/t$; hence

$$\frac{H}{x^2}\sim\frac{4\pi\sigma\mu}{c^2}\frac{H}{t}, \qquad \text{i.e.} \qquad \frac{x}{t}\sim\frac{c^2}{4\pi\sigma\mu x}.$$

This, roughly speaking, is the velocity at which the magnetic field penetrates the conductor. Thus if equation (1) is valid, the magnetic field does not diffuse through the material.

The rate of flow of momentum (per unit area) associated with the flow velocity is given roughly by

$$\rho v^2\sim B^2/8\pi\mu \qquad \ldots\ldots(2)$$

because it results from the inward-directed magnetic force. Hence condition (1) can be written

$$\frac{Bx\sigma}{c^2}\left(\frac{2\pi\mu}{\rho}\right)^{1/2}\gg 1. \qquad \ldots\ldots(3)$$

Lundquist (1952) has given a similar criterion for magnetohydrodynamic coupling (his expression differs from (3) only by a factor of $\sqrt{2}$).

Condition (*b*).

It has been assumed in §2 that the current rises instantaneously to a finite value and then remains constant. To approach this condition in practice the current must be established in a time which is short compared with the shock-wave transit time t_1, and must then remain sensibly constant during the latter. It can be shown, using the results in table 1 together with the relation $p_B=B^2/8\pi\mu$, that

$$t_1=\frac{cd}{2I}\left(\frac{3\rho_A}{2\pi\mu}\right)^{1/2} \qquad \ldots\ldots(4)$$

for strong shocks, where d is the width of the discharge chamber and I/c is the total current in e.m.u. ($I=2I_3$).

The time to maximum contraction of the plasma (t_2 in figure 2) is 20% greater than t_1 in the case of a strong shock.

Condition (*c*).

The shock wave formulae presented in tables 1 and 2 are valid only if the shock wave thickness is small compared with the dimensions of the discharge chamber. Under these conditions it can easily be shown that the formulae remain valid for shock waves in an ionized gas provided that the 'temperature' T is defined by the relation

$$T=(Z+1)T_1 \qquad \ldots\ldots(5)$$

where Z is the ionic charge and T_1 is the kinetic temperature of the electrons, equal to T_2, the kinetic temperature of the ions. The effective 'molecular weight' of the gas is unaltered, i.e.

$$m=Zm_1+m_2 \qquad \ldots\ldots(6)$$

where m_1 and m_2 are the masses of the electron and ion respectively.

In this connection the term 'shock wave thickness' refers to the total width of the transition zone between the high and low pressure gases. Marshall (1956) has studied theoretically the structure of shock waves in a conducting fluid, but no satisfactory theory of the shock structure in an ionized gas has yet been published. Such a theory must take into account the temperature difference between the electrons and ions, in the transition zone, and also the long range (space charge) forces due to charge separation. Denisse and Rocard (1951) have considered the latter (but not the former), and measurements of the potential difference across the shock front have been made by Petschek (1955).

The transition zone in an ionized gas consists of a relatively sharp 'front' of width of the order of a mean free path in which the ion temperature rises appreciably, followed by a long 'tail' in which equilibrium between electrons and ions is approached. In both regions the electrons and ions interact via the long range (space charge) forces in addition to the short range (collision) forces. The latter are assumed to play the major role in the 'tail' in the following rough calculation of the width of the transition zone. A high energy ion loses only a small fraction ($\sim 2m_1/m_2$) of its energy in a collision with a slow electron, so that an electron makes about $m_2/2m_1$ collisions, on the average, before it approaches equilibrium with the ions. During this approach to equilibrium, the motion of an electron approximates to a three-dimensional random walk, so that the length of the 'tail', i.e. the overall width δ of the transition zone is given by

$$\delta^2 \sim \tfrac{1}{3}N\lambda^2 \text{ where } N = m_2/2m_1 \text{ and } \lambda \text{ is the mean free path.} \quad \ldots\ldots(7)$$

Thus δ is about 17λ for ionized hydrogen and 25λ for ionized deuterium.

§ 4. DISCUSSION

The shock waves described in § 2 are large-scale plasma ion oscillations (of the acoustic type). Electrostatic fields play an essential part in the mechanism of these oscillations, even though electrical neutrality is preserved to a high degree of approximation (Tonks and Langmuir 1929).

A theory of the collapse of a high current discharge plasma, due to its self-magnetic pinch force, has also been developed by Leontovich and Osovets (1956). Their theory differs from the present treatment in that shock phenomena are not considered; instead, it is assumed that the gas undergoes an adiabatic compression. This latter assumption, however, has little effect on their calculated contraction times, which are determined essentially by the inertia of the gas.

The theory described in § 2 does not refer to discharges at very low pressures, because shock wave formation does not take place if the ionic mean free path is greater than the width of the discharge chamber. Neither does it apply to discharges at very high pressures (of the order of atmospheric) where a different behaviour is observed (Allen and Craggs 1954), mainly because the initial breakdown of the gas is localized. In practice, ionizing and other inelastic collisions occur, the electrical conductivity is not always high enough to satisfy equation (3), and complex current waveforms are recorded. Also, experimental discharge plasmas are usually cylindrical, whereas a plane geometry has been considered here. Nevertheless, the model described in § 2 is of interest, since it provides an explanation of the shock wave phenomena which were first observed by Cousins and Ware (1951) and which have been recorded in this laboratory.

Acknowledgments

The author wishes to acknowledge the interest shown in this work by Dr. P. C. Thonemann and Mr. R. S. Pease. The development of the theory resulted from a study of rotating-mirror photographs of high current discharges taken by Dr. P. Reynolds, Mr. L. S. Holmes and the author.

References

Allen, J. E., and Craggs, J. D., 1954, *Brit. J. Appl. Phys.*, **5**, 446.
Artsimovich, L. A., Andrianov, A. N., Bazilevskaya, O. A., Prokhorov, Yu.G., and Filippov, N. V., 1956, *Atomnaya Energiya*, **3**, 76.
Courant, R., and Friedrichs, K. O., 1948, *Supersonic Flow and Shock Waves* (New York: Interscience).
Cousins, S. W., and Ware, A. A., 1951, *Proc. Phys. Soc.* B, **64**, 159.
Denisse, J. F., and Rocard, Y., 1951, *J. Phys. Radium*, **12**, 893.
Leontovich, M. A., and Osovets, S. M., 1956, *Atomnaya Energiya*, **3**, 81.
Lundquist, S., 1952, *Ark. Fys.*, **5**, 297.
Marshall, W., 1956, *Proc. Roy. Soc.* A, **233**, 367.
Petschek, H. E., 1955, *Thesis*, Cornell University.
Tonks, L., and Langmuir, I., 1929, *Phys. Rev.*, **33**, 195, 990.

Hydromagnetic Instabilities of an Ideally Conducting Fluid

BY R. J. TAYLER
Atomic Energy Research Establishment, Harwell, Berks.

Communicated by B. H. Flowers; MS. received 8th November 1956

Abstract. Some hydromagnetic instabilities of a cylindrical ideally conducting fluid, which is confined by the pinch effect of an axial current, are considered. Kruskal and Schwarzschild have previously studied the instabilities of a fluid carrying a surface current. We consider a volume current distributed according to a power law, and some general stability criteria and values of maximum possible growth rates are obtained. The volume current is more stable than the surface current, and in particular an infinite instability at short wavelength is replaced by one of finite growth rate. Finally, viscosity is introduced into the plasma carrying a surface current; disturbances at very short wavelength are damped, and a first approximation is obtained to a wavelength of maximum instability.

§ 1. INTRODUCTION

A HIGHLY constricted gas discharge is observed to be unstable against disturbances which distort the shape of its surface. In this paper an attempt is made towards a theoretical understanding of these hydromagnetic instabilities of the pinch effect. Previous work on this problem includes that of Kruskal and Schwarzschild (1954); we generally follow their notation except that we have used gaussian rather than m.k.s. units for the electrical quantities.

The problem treated here is highly simplified for mathematical reasons. In equilibrium the current carrying region is restricted to a cylinder and has infinite electrical conductivity. This is surrounded by a medium of zero conductivity. In the model studied by Kruskal and Schwarzschild the current causing the pinch is confined to an extremely thin sheet on the surface of the conducting fluid; we shall generalize this to consider the case of a fluid carrying a current whose density at any point is proportional to $(r/r_0)^n$. Our object is to obtain stability criteria in terms of the exponent n and an azimuthal wave number m, and in the case where instability occurs to obtain upper bounds to the possible instability growth rates. One particular result of interest is the infinite growth rate at zero wavelength found by Kruskal and Schwarzschild; it is found that if the current is distributed through the plasma the infinity is removed, and an expression for the maximum growth rate in terms of n can be obtained.

Although our main interest is in the instabilities of a compressible plasma, most of the work is carried though for both compressible and incompressible fluids. In the case of an incompressible fluid of constant density the equations are considerably simplified, and stability criteria can be obtained without an excessive amount of algebra; afterwards the work is repeated for a compressible plasma and the general stability criteria are found to be unaltered although maximum instability growth rates are changed.

All the work of this paper is based on the hydromagnetic equations and is thus macroscopic in character. The fluid is taken to be ideally conducting and, in what we have described above, we have neglected other mean free path phenomena such as viscosity and heat conduction. In §§ 7 and 8, by introducing viscous terms into the equation of motion, we attempt to take some account of the finite mean free path while remaining within the framework of hydrodynamics. When results are obtained we must try to decide whether those of most interest occur in conditions in which the macroscopic equations apply.

Our object in this paper is to obtain what interesting results we can without being involved in extremely long calculations. Thus we obtain sufficiency criteria for stability and upper bounds for instability growth rates without actually solving the hydromagnetic equations. In the cases in which a complete solution of these equations has been obtained, we have been able to express the solutions in terms of Bessel functions only. If actual values of instability growth rates were required instead of upper bounds, an automatic computer would probably have to be used.

The remainder of the paper is arranged as follows: in § 2 the basic equations and method of solution are discussed, in §3 the instabilities of a sheet current are studied, in §§ 4, 5 and 6 we consider the instabilities of a volume current, in §§ 7 and 8 viscosity is introduced, and our results are discussed in § 9.

§ 2. BASIC EQUATIONS AND METHOD OF SOLUTION

The hydromagnetic equations for a medium of pressure, density, velocity, and ratio of specific heats, p, ρ, $\mathbf{v}$ and γ respectively, which carries electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$ and current and charge densities $\mathbf{j}$ and ϵ, will take the following form.

Maxwell's equations:

$$\operatorname{curl}\mathbf{B}=\frac{4\pi\mathbf{j}}{c}, \qquad \ldots\ldots(2.1)$$

$$\operatorname{curl}\mathbf{E}=-\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t} \qquad \ldots\ldots(2.2)$$

$$\operatorname{div}\mathbf{B}=0, \qquad \ldots\ldots(2.3)$$

$$\operatorname{div}\mathbf{E}=4\pi\epsilon. \qquad \ldots\ldots(2.4)$$

The equations are expressed in terms of gaussian units and displacement currents are neglected throughout the present paper.

Equation of motion:

$$\rho\frac{d\mathbf{v}}{dt}=-\operatorname{grad}p+\frac{\mathbf{j}\times\mathbf{B}}{c}. \qquad \ldots\ldots(2.5)$$

In the last problem of the paper, viscosity is introduced in the ideally conducting fluid, but at the same time the fluid bears no electromagnetic fields. Equation (2.5) is replaced by

$$\rho\frac{d\mathbf{v}}{dt}=-\operatorname{grad}p+\frac{4\mu}{3}\operatorname{grad}\operatorname{div}\mathbf{v}-\mu\operatorname{curl}\operatorname{curl}\mathbf{v}, \qquad \ldots\ldots(2.6)$$

where μ is the coefficient of viscosity.

Equation of mass conservation:

$$\frac{\partial\rho}{\partial t}=-\operatorname{div}\rho\mathbf{v}. \qquad \ldots\ldots(2.7)$$

In the compressible plasma we assume an adiabatic law:

$$\frac{1}{p}\frac{dp}{dt}=\frac{\gamma}{\rho}\frac{d\rho}{dt}, \qquad \ldots\ldots(2.8)$$

thus neglecting heating and heat losses by conduction. In the incompressible fluid equation (2.8) is replaced by

$$\operatorname{div}\mathbf{v}=0. \qquad \ldots\ldots(2.9)$$

In a medium of perfect electrical conductivity we have

$$\mathbf{E}+\frac{\mathbf{v}\times\mathbf{B}}{c}=0, \qquad \ldots\ldots(2.10)$$

where the Hall effect and other cross terms have been neglected.

In a medium of zero conductivity

$$\mathbf{j}=0. \qquad \ldots\ldots(2.11)$$

Boundary conditions at interface between two media:

These conditions have been given in full by Kruskal and Schwarzschild (1954). At the interface between two media there may be sheet currents and sheet charges $\mathbf{j}^*$ and ϵ^* and discontinuities in the other physical variables. If $\mathbf{n}$ is the unit normal to the surface directed into the perfect conductor, and the surface has normal velocity $u\mathbf{n}$, the surface conditions are

$$\mathbf{n}\times[\mathbf{B}]=\frac{4\pi}{c}\mathbf{j}^*, \qquad \ldots\ldots(2.12)$$

$$\mathbf{n}\,.\,[\mathbf{B}]=0, \qquad \ldots\ldots(2.13)$$

$$\mathbf{n}\times[\mathbf{E}]=\frac{u}{c}[\mathbf{B}], \qquad \ldots\ldots(2.14)$$

$$\mathbf{n}\,.\,[\mathbf{E}]=4\pi\epsilon^* \qquad \ldots\ldots(2.15)$$

and

$$(\mathbf{j}^*\times\overline{\mathbf{B}})_\beta=-cn_\alpha p_{\alpha\beta}. \qquad \ldots\ldots(2.16)$$

Brackets around any quantity denote the difference between the values of that quantity in the conducting and non-conducting media, and $\overline{\mathbf{B}}$ denotes the mean value of $\mathbf{B}$ at the interface. In equation (2.16) $p_{\alpha\beta}$ is the mechanical stress tensor which in cylindrical coordinates has components

$$p_{rr}=-\left(p+\frac{2\mu}{3}\operatorname{div}\mathbf{v}\right)+2\mu\frac{\partial v_r}{\partial r}, \qquad \ldots\ldots(2.17)$$

$$p_{\theta\theta}=-\left(p+\frac{2\mu}{3}\operatorname{div}\mathbf{v}\right)+2\mu\left(\frac{1}{r}\frac{\partial v_\theta}{\partial\theta}+\frac{v_r}{r}\right), \qquad \ldots\ldots(2.18)$$

$$p_{zz}=-\left(p+\frac{2\mu}{3}\operatorname{div}\mathbf{v}\right)+2\mu\frac{\partial v_z}{\partial z}, \qquad \ldots\ldots(2.19)$$

$$p_{\theta z}=\mu\left(\frac{1}{r}\frac{\partial v_z}{\partial\theta}+\frac{\partial v_\theta}{\partial z}\right)=p_{z\theta}, \qquad \ldots\ldots(2.20)$$

$$p_{zr}=\mu\left(\frac{\partial v_r}{\partial z}+\frac{\partial v_z}{\partial r}\right)=p_{rz} \qquad \ldots\ldots(2.21)$$

and

$$p_{r\theta}=\mu\left(\frac{1}{r}\frac{\partial v_r}{\partial\theta}-\frac{v_\theta}{r}+\frac{\partial v_\theta}{\partial r}\right)=p_{\theta r}. \qquad \ldots\ldots(2.22)$$

In addition the normal velocity and unit normal vector are related by

$$u=\mathbf{n}\,.\,\mathbf{v} \qquad \ldots\ldots(2.23)$$

and

$$\frac{d\mathbf{n}}{dt}=\mathbf{n}\times(\mathbf{n}\times\operatorname{grad}u). \qquad \ldots\ldots(2.24)$$

In any particular problem it may not be necessary to apply the entire set of equations (2.1) to (2.24) to obtain the stability properties. Some equations are required only if we need values of all variables in a perturbed configuration.

In all the problems we consider we shall have a cylinder of ideally conducting fluid carrying either a sheet current or a volume current. It will be in equilibrium surrounded by a medium of vanishing electrical conductivity. We shall then consider perturbations of such an equilibrium configuration in which any variable q takes the form

$$q=q_e+\tilde{q}=q_e+\hat{q}(r)\exp\{i(m\theta+kz)+\omega t\}, \qquad \ldots\ldots(2.25)$$

where q_e is the equilibrium value of q, k is a real wave number, m is an integer (positive or negative) and ω is to be determined. The general perturbation will consist of a sum over m and integral over k of terms of the above form, but as we linearize our equations each mode can be treated independently. The relevant linearized equations from the set (2.1) to (2.11) can be solved in each medium, and expressions can be obtained for all the perturbed quantities in terms of constants which are for the time being arbitrary. When the conditions (2.12) to (2.16) are applied, relations are found between the constants, and on eliminating the constants we obtain a dispersion relation between m and k and the disturbance growth rate ω. Unstable modes will correspond to values of m and k for which the dispersion relation has a root for ω with a positive real part. In this paper we assume that any instability occurs through 'convection', rather than 'over-stability'. This means that any instability is associated with a real positive value of ω. In the particular problem of the sheet current studied in the next section, it is easy to show that over-stability can never occur; the proof, though simple, is lengthy and will be omitted.

As the equations applicable in the outer medium will not change from one problem to the next, we will complete this section by considering the solution of these equations. As the electrical conductivity vanishes the electromagnetism and hydrodynamics are not linked. It is easy to show that the perturbed quantities satisfy the equations

$$rD(r\tilde{p})-\left[\left(k^2+\frac{\rho_1\omega^2}{\gamma_1 p_1}\right)r^2+m^2\right]\tilde{p}=0 \qquad \ldots\ldots(2.26)$$

and

$$rD(r\tilde{B}_z)-[k^2r^2+m^2]\tilde{B}_z=0, \qquad \ldots\ldots(2.27)$$

where D is d/dr, and p_1, ρ_1, and γ_1 are the equilibrium pressure, density and ratio of specific heats of the outer medium.

Then we can write

$$\tilde{p}^{\mathrm{O}}=\tilde{p}_0^{\mathrm{O}}\ \mathrm{K}_{|m|}(\zeta_1 r)\exp\{i(m\theta+kz)+\omega t\}, \qquad \ldots\ldots(2.28)$$

$$\tilde{v}_r^{\mathrm{O}}=-\frac{\zeta_1}{\rho_1\omega}\tilde{p}_0^{\mathrm{O}}\,\mathrm{K}'_{|m|}(\zeta_1 r)\exp\{i(m\theta+kz)+\omega t\}, \qquad \ldots\ldots(2.29)$$

$$\tilde{B}_z^{\mathrm{O}}=\tilde{B}_{z0}^{\mathrm{O}}\,\mathrm{K}_{|m|}(|k|r)\exp\{i(m\theta+kz)+\omega t\}, \qquad \ldots\ldots(2.30)$$

$$\tilde{B}_r^{\mathrm{O}}=-\frac{ik}{|k|}\tilde{B}_{z0}^{\mathrm{O}}\,\mathrm{K}'_{|m|}(|k|r)\exp\{i(m\theta+kz)+\omega t\}, \qquad \ldots(2.31)$$

and

$$\tilde{B}_\theta^{\mathrm{O}}=\frac{m}{kr}\tilde{B}_{z0}^{\mathrm{O}}\,\mathrm{K}_{|m|}(|k|r)\exp\{i(m\theta+kz)+\omega t\}, \qquad \ldots\ldots(2.32)$$

where $\tilde{p}_0^{\mathrm{O}}$, $\tilde{B}_{z0}^{\mathrm{O}}$ are constants and the superscript O refers to the outer medium. ζ_1 is defined by $\zeta_1^2=k^2+\rho_1\omega^2/\gamma_1 p_1$, and ζ_1 must be taken to be positive or to have a positive real part. The perturbed quantities given above are the only ones required in obtaining the dispersion relation.

§ 3. INSTABILITIES OF A SHEET CURRENT

The model considered here will be essentially one discussed by Kruskal and Schwarzschild (1954). They, however, only considered perturbations of the form $\exp\{i(\theta+kz)+\omega t\}$; we consider the slightly more general perturbation $\exp\{i(m\theta+kz)+\omega t\}$.

A perfectly conducting fluid of uniform density ρ_0, pressure p_0 and ratio of specific heats γ, forms a cylinder of radius r_0. The fluid bears a surface current $(0,0,j_0{}^*)$, and is surrounded by a vacuum in which there is a magnetic field $(0, B_{\theta 0}r_0/r, 0)$. This system is in equilibrium provided that

$$B_{\theta 0}=4\pi j_0{}^*/c \text{ and } B_{\theta 0}{}^2/8\pi=p_0. \qquad \ldots\ldots(3.1)$$

In the previous section we have considered the solution of the perturbed vacuum equations (we put $\tilde{p}_0{}^{\mathrm{O}}=0$ for a vacuum). In the plasma all the perturbed qualities can be expressed in terms of $\tilde{p}$, where $\tilde{p}$ satisfies the equation

$$rD[rD\tilde{p}]-\left[\left(k^2+\frac{\omega^2}{c_s{}^2}\right)r^2+m^2\right]\tilde{p}=0, \qquad \ldots\ldots(3.2)$$

where D is d/dr and $c_s{}^2=\dfrac{\gamma p_0}{\rho_0}$ (c_s=velocity of sound).

We write $k^2+\omega^2/c_s{}^2=\zeta^2$ where ζ is to have a positive real part.

Then $$\rho=\tilde{p}_0\mathrm{I}_{|m|}(\zeta r)\exp\{i(m\theta+kz)+\omega t\}. \qquad \ldots\ldots(3.3)$$

The perturbed density and velocity components are given by

$$\tilde{\rho}=\frac{1}{c_s{}^2}\tilde{p}_0\,\mathrm{I}_{|m|}(\zeta r), \qquad \ldots\ldots(3.4)$$

$$\tilde{v}_r=-\frac{\zeta}{\rho_0\omega}\tilde{p}_0\,\mathrm{I}'_{|m|}(\zeta r), \qquad \ldots\ldots(3.5)$$

$$\tilde{v}_\theta=-\frac{im}{\rho_0\omega r}\tilde{p}_0\,\mathrm{I}_{|m|}(\zeta r), \qquad \ldots\ldots(3.6)$$

$$\tilde{v}_z=-\frac{ik}{\rho_0\omega}\tilde{p}_0\,\mathrm{I}_{|m|}(\zeta r), \qquad \ldots\ldots(3.7)$$

where the exponential factor has been omitted from the right-hand side of equations (3.4) to (3.7).

We now apply the boundary conditions (2.12) to (2.16) at the plasma surface and obtain the dispersion relation

$$\frac{\gamma X_0 W_0{}^2}{2(X_0{}^2+W_0{}^2)^{1/2}}\,\frac{\mathrm{I}_{|m|}[(X_0{}^2+W_0{}^2)^{1/2}]}{\mathrm{I}'_{|m|}[(X_0{}^2+W_0{}^2)^{1/2}]}=X_0+m^2\frac{\mathrm{K}_{|m|}(X_0)}{\mathrm{K}'_{|m|}(X_0)}, \qquad \ldots\ldots(3.8)$$

where we have introduced the non-dimensional variables $X_0=|k|r_0$ and $W_0{}^2=\omega^2 r_0{}^2/c_s{}^2$. Equation (3.8) is equivalent to equation (31) of Kruskal and Schwarzschild when $m=1$. As this equation is even in m we take m to be positive in what follows.

As mentioned in the previous section, it can be proved that there are no over-stable solutions to this dispersion relation; in addition the left-hand side of equation (3.8) is positive if $W_0^2>0$, and varies from 0 to ∞ as W_0^2 varies from 0 to ∞. Instability thus occurs for values of X_0 and m such that

$$X_0+m^2\frac{K_{|m|}(X_0)}{K'_{|m|}(X_0)}>0. \qquad \ldots\ldots(3.9)$$

For $m=0$ and $m=1$ the plasma is unstable for all wave numbers; for higher values of m there is a range of wavenumbers extending from $X_0=0$ for which we obtain stability. For large enough X_0, $K_m(X_0)/K_m'(X_0)\to -1$, and the largest value of X_0 for which stability occurs is approximately m^2. Table 1 gives results for different m; there is stability when $X_0<(X_0)_{\text{crit}}$.

Table 1

m	2	3	4	5
$(X_0)_{\text{crit}}$	3·04	8·02	15·01	24·01

For small X_0 there is only instability for $m=0$ and 1. For the case $m=1$, the growth rate ω is given by

$$\omega\simeq|k|(2Lp_0/\rho_0)^{1/2}, \qquad \ldots\ldots(3.10)$$

where $L=\ln(2/X_0)-C+1$. (C is Euler's constant.)

There is instability for all m for sufficiently short wavelengths, and in the limit the same growth rate is obtained for all m:

$$\omega\simeq(2|k|p_0/\rho_0 r_0)^{1/2}. \qquad \ldots\ldots(3.11)$$

These results (3.10), (3.11) were given by Kruskal and Schwarzschild (1954). Equation (3.11) predicts that disturbances of any azimuthal wave number m will have an infinite growth rate at zero wavelength. We cannot expect this formula to hold at arbitrarily small wavelengths because of the eventual failure of the macroscopic equations; however, the present theory would suggest that the shortest wavelength disturbances should be the most unstable. The main concern of the remainder of this paper is with the instability at short wavelengths.

In the next three sections of this paper we consider the instabilities of a conducting fluid carrying a current which is no longer confined to its surface, but which is proportional to $(r/r_0)^n$. When $n=0$ we have a uniform volume current; as $n\to\infty$ we approach the case studied above. Before considering the general case of a compressible fluid it is convenient to study the instabilities of an incompressible fluid of constant density; the equations are simpler, and there exists one problem, involving an incompressible fluid and a volume current, for which a solution can be obtained in terms of simple functions. Furthermore, it will be seen later that the results for incompressible and compressible fluids are basically similar. Indeed, at one time it seemed that it could be proved that conditions for stability did not depend on the density or compressibility of the fluid, since the equations with the growth rate ω equal to zero do not depend on these two quantities (Dungey and Loughhead 1954, Loughhead 1955). However, as we have pointed out elsewhere (Tayler 1957), this proof breaks down in a problem such as this, when more than one growth rate can be associated with a given wavenumber.

We complete this section by observing that for an incompressible fluid equation (3.8) takes the form

$$Y_0^2\frac{I_m(X_0)}{I_m'(X_0)}=X_0+m^2\frac{K_m(X_0)}{K_m'(X_0)}, \qquad \ldots\ldots(3.12)$$

where $Y_0^2=4\pi\rho_0\omega^2r_0^2/B_{\theta 0}^2$. The instability at short wavelengths is still given by $\omega\simeq(2|k|p_0/\rho_0 r_0)^{1/2}$.

§ 4. INSTABILITIES OF INCOMPRESSIBLE FLUID CARRYING A UNIFORM CURRENT

In the following two sections we consider the instabilities of incompressible and compressible fluids carrying axial currents whose density is proportional to $(r/r_0)^n$ with $n \geqslant 0$. The hydrodynamic equations in the conducting fluid cannot in general be solved in terms of simple functions but we are able to deduce general stability properties without actually solving these equations. In this section we consider the solution of one particular problem involving only simple functions, for which full results have been computed.

In equilibrium we consider a cylinder of radius r_0 of an incompressible ideally conducting fluid of uniform density ρ_0 and pressure $p_0\,[1-(r/r_1)^2]$. The fluid carries a uniform current $j_z = j_0$ and a magnetic field $B_\theta = B_{\theta 0}r/r_0$, where

$$p_0 = \frac{B_{\theta 0}^2}{4\pi}\left(\frac{r_1}{r_0}\right)^2 \text{ and } j_0 = \frac{cB_\theta}{2\pi r_0}. \qquad \ldots\ldots(4.1)$$

Surrounding the conducting fluid is another incompressible medium of density ρ_1, pressure $p_1 = p_0[1-(r_0/r_1)^2]$ and carrying a magnetic field $B_\theta = B_{\theta 0}r_0/r$. In this section and the following two an outer medium with non-vanishing density and pressure is needed to avoid an unsatisfactory feature in the solution of the problem. If we put $r_0 = r_1$, and $p_1 = \rho_1 = 0$, then the equilibrium pressure vanishes on $r = r_0$, and after a perturbation has been applied negative pressures occur on the fluid boundary. As we are only considering a linear stability problem, the introduction of a medium of any density however small will remove this trouble. For the fluid of constant density the presence of the outer medium has no great effect on the solution of the stability problem and, in fact, calculations have been made for the case of a vacuum. In the case considered in § 6 the value of the ratio of r_0 to r_1 may affect the magnitude of instabilities.

We now consider perturbations of the equilibrium configuration. Equations (2.2), (2.9) and (2.10) combine to give $\tilde{\mathbf{B}} = (imB_{\theta 0}/\omega r_0)\tilde{\mathbf{v}}$. Then the second component of equation (2.5) gives

$$p = \frac{i\rho_0\omega r\tilde{v}_\theta}{m} + \frac{B_{\theta 0}^2}{2\pi\omega r_0^2}r\tilde{v}_r. \qquad \ldots\ldots(4.2)$$

The other two components of equation (2.5) combined with equations (4.2) and (2.9) lead to two equations for $\tilde{v}_r$ and $\tilde{v}_\theta$:

$$2mD(r\tilde{v}_r) + m\tilde{v}_r[Y_0^2 + m^2] + iD(r\tilde{v}_\theta)[Y_0^2 + m^2] + 2im^2\tilde{v}_\theta = 0$$

and

$$m[Y_0^2 + m^2]D(r\tilde{v}_r) + 2k^2r^2m\tilde{v}_r + i\tilde{v}_\theta[Y_0^2 + m^2][m^2 + k^2r^2] = 0$$

where $Y_0^2 = 4\pi\rho_0\omega^2r_0^2/B_{\theta 0}^2$. Now put $m\tilde{v}_r = \phi$, $\beta\phi + i\tilde{v}_\theta = \chi$, where $\beta = 2/(Y_0^2 + m^2)$.

Then

$$D(r\chi) + m^2\beta\chi + \phi[1 - m^2\beta^2] = 0 \qquad \ldots\ldots(4.3)$$

and

$$D(r\phi) + \chi(m^2 + k^2r^2) - m^2\beta\phi = 0. \qquad \ldots\ldots(4.4)$$

If ϕ is now eliminated

$$D[rD(r\chi)] + [(m^2\beta^2 - 1)k^2r^2 - m^2]\chi = 0, \qquad \ldots\ldots(4.5)$$

leading to

$$\chi \propto \mathrm{J}_{|m|}\{(m^2\beta^2 - 1)^{1/2}|k|r\}/|k|r \qquad \ldots\ldots(4.6)$$

and

$$\phi \propto \frac{m^2\beta\,\mathrm{J}_{|m|}\{(m^2\beta^2 - 1)^{1/2}|k|r\}}{(m^2\beta^2 - 1)|k|r} + \frac{\mathrm{J}_{|m|}'\{(m^2\beta^2 - 1)^{1/2}|k|r\}}{(m^2\beta^2 - 1)^{1/2}}. \qquad \ldots\ldots(4.7)$$

When the boundary conditions (2.12) to (2.16) are applied the dispersion relation can be obtained in the form

$$\left(\frac{\phi}{\chi}\right)_{r=r_0} = -\frac{m^2 + Y_0^2}{m^2 + Y_0^2\rho_1/\rho_0}|k|r_0\frac{\mathrm{K}_{|m|}'(|k|r_0)}{\mathrm{K}_{|m|}(|k|r_0)}. \qquad \ldots\ldots(4.8)$$

As equations (4.6) to (4.8) are even in m and k we may drop the modulus sign and assume $m, k>0$.

When $m^2\beta^2<1$ the Bessel functions become modified Bessel functions and it can then easily be shown that the two sides of equation (4.8) have opposite signs if $Y_0^2>0$. Thus for position Y_0^2, $m^2\beta^2>1$. However, since $\beta=2/(Y_0^2+m^2)$ it can be seen that if $m>1$, $m^2\beta^2<1$ and thus the dispersion relation can have no positive roots for Y_0^2. Thus there are no unstable modes with m greater than 1. This immediately shows a great increase in stability compared with the problem of the sheet current considered in the previous section where instabilities exist for all m. In the case $m=1$, the inequality connecting m and β shows that solutions can only occur with Y_0^2 less than 1. The dispersion relation for the case $m=1$ has been solved numerically by Mr. E. B. Fossey of the Computing Group of A.E.R.E., Harwell, and the results are shown in table 2 and graphically in figure 1. The calculation is not completely straightforward as for any one wave-number k the dispersion relation may have many positive roots for Y_0^2. In fact, the largest such root is required. (A similar problem is considered by Tayler (1957).)

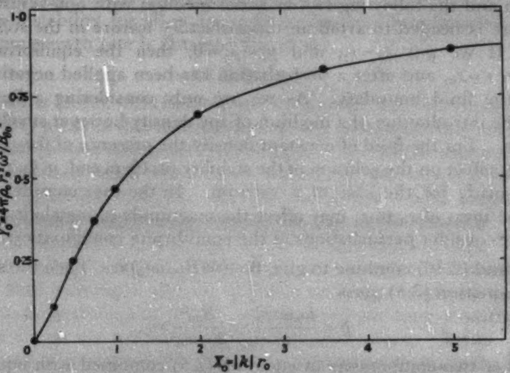


Figure 1. Plot of non-dimensional growth rate Y_0^2 against non-dimensional wave number X_0 for the unstable $m=1$ mode of an incompressible fluid carrying a current of uniform density.

Table 2

$\lvert k\rvert r_0$	0·00	0·25	0·50	0·75	1·00	2·00	3·50	5·00	10·00	∞
Y_0^2	0·0000	0·1048	0·2481	0·3702	0·4670	0·6935	0·8319	0·8931	0·9616	1·0000

It can be seen from table 2 that the maximum growth rate still occurs at zero wavelength, but that this growth rate is no longer infinite.

§ 5. INCOMPRESSIBLE FLUID: GENERAL CURRENT DISTRIBUTION

We consider now a more general case than that studied in the last section. The conducting fluid carries an axial current proportional to r^n, and we consider perturbations in which all variables are proportional to $\exp\{i(m\theta+kz)+\omega t\}$. We find that it is possible to obtain sufficiency conditions for stability against given perturbations in the form of inequalities involving m and n. In cases where instability occurs we can place upper bounds on the possible instability growth rates.

In equilibrium we have a cylindrical conducting fluid of radius r_0. The equilibrium values of the physical variables are

$$\rho=\rho_0,\ p=p_0[1-(r/r_1)^{2n+2}],$$

$$\mathbf{B}=[0,\ B_{\theta0}(r/r_0)^{n+1},\ 0],\ \mathbf{j}=\left[0,0,\frac{(n+2)cB_{\theta0}}{4\pi r_0}\left(\frac{r}{r_0}\right)^n\right],$$

where

$$\frac{B_{00}{}^2}{4\pi}\frac{n+2}{2n+2}\left(\frac{r_1}{r_0}\right)^{2n+2}=p_0.$$

The conducting fluid is surrounded by a medium of density ρ_1, pressure $p_1=p_0[1-(r_0/r_1)^{2n+2}]$, and carrying a magnetic field $B_\theta=B_{\theta0}r_0/r$. This medium is also incompressible.

We now consider the solution for the perturbed quantities in the conducting fluid. Equations (2.2), (2.9) and (2.10) combine to give

$$\tilde{\mathbf{B}}=\frac{iB_{\theta0}}{\omega r_0}\left(\frac{r}{r_0}\right)^n[m\tilde{v}_r, m\tilde{v}_\theta+in\tilde{v}_r, m\tilde{v}_z]. \qquad \ldots\ldots(5.1)$$

The second component of equation (2.5) leads to the equation

$$\tilde{p}=\frac{i\rho_0\omega}{m}(r\tilde{v}_\theta)+\frac{(n+2)B_{\theta0}{}^2}{4\pi\omega r_0{}^2}\left(\frac{r}{r_0}\right)^{2n}r\tilde{v}_r. \qquad \ldots\ldots(5.2)$$

Using equations (2.9) and (5.2) in conjunction with the other two components of equation (2.5) we obtain two equations for $\tilde{v}_r$ and $\tilde{v}_\theta$:

$$2m\left(\frac{r}{r_0}\right)^{2n}D(r\tilde{v}_r)+m\tilde{v}_r\left[Y_0{}^2+(2n+m^2)\left(\frac{r}{r_0}\right)^{2n}\right]$$
$$+iD(r\tilde{v}_\theta)\left[Y_0{}^2+m^2\left(\frac{r}{r_0}\right)^{2n}\right]+i\tilde{v}_\theta\left[(2n+2)m^2\left(\frac{r}{r_0}\right)^{2n}\right]=0 \qquad \ldots\ldots(5.3)$$

and

$$mD(r\tilde{v}_r)\left[Y_0{}^2+m^2\left(\frac{r}{r_0}\right)^{2n}\right]+2mk^2r^2\tilde{v}_r+i\tilde{v}_\theta[m^2+k^2r^2]\left[Y_0{}^2+m^2\left(\frac{r}{r_0}\right)^{2n}\right]=0, \qquad \ldots\ldots(5.4)$$

where $Y_0{}^2=4\pi\rho_0\omega^2r_0{}^2/B_{\theta0}{}^2$. We now make the further substitutions $m\tilde{v}_r=\phi$, $\beta m\tilde{v}_r+i\tilde{v}_\theta=\chi$, where

$$\beta=2\left(\frac{r}{r}\right)^{2n}\bigg/\left[Y_0{}^2+m^2\left(\frac{r}{r_0}\right)^{2n}\right].$$

Then equations (5.3) and (5.4) become

$$D(r\chi)+\chi(n+1)m^2\beta+\phi[1-n\beta-m^2\beta^2]=0 \qquad \ldots\ldots(5.5)$$

and

$$D(r\phi)+\chi(m^2+k^2r^2)-\phi m^2\beta=0. \qquad \ldots\ldots(5.6)$$

Let us now consider the boundary conditions at the surface of the perfect conductor. When these are applied we obtain the dispersion relation between ω, k and m in the form

$$\frac{\phi}{\chi}=-\frac{[m^2+Y_0{}^2]}{[m^2+Y_0{}^2\rho_1/\rho_0]}|k|r_0\frac{\mathrm{K}_{|m|}{}'(|k|r_0)}{\mathrm{K}_{|m|}(|k|r_0)}, \qquad \ldots\ldots(5.7)$$

where the functions ϕ and χ are to be evaluated at $r=r_0$.

We may note that equations (5.5), (5.6) and (5.7) are all even in m and k and thus in obtaining stability criteria we can take $m \geqslant 0$, $k \geqslant 0$. In addition the derivation of equations (5.5) to (5.7) given here implicitly assumes $m \neq 0$; thus at present we restrict ourselves to $m \geqslant 1$.

It can be seen that the right-hand side of equation (5.7) is always positive if $Y_0^2 > 0$ (positive growth rate); thus solutions of the dispersion relation with positive growth rate can only occur when $\phi(r_0)$ and $\chi(r_0)$ have the same sign.

Consider now equations (5.5) and (5.6) near $r=0$. First if $n>0$ we can show that, near $r=0$, $\phi = ar^{m-1}$, $\chi = br^{m-1}$, where $a = -mb$. Thus, if $n>0$, ϕ and χ have opposite signs when $r=r_0$. Furthermore, we can show that if $n=0$, ϕ and χ have opposite signs when $r=0$ if $m=1$ and $Y_0^2>1$ or $m>1$ and $Y_0^2>0$. Thus, omitting the case $m=1$, $n=0$ which we have previously studied, ϕ and χ have opposite signs at $r=0$ when $Y_0^2>0$.

Solutions of the dispersion relation (with $Y_0^2>0$) can, therefore, only occur if ϕ or χ changes sign between $r=0$ and $r=r_0$. It can be seen simply from equation (5.6) that ϕ cannot have a zero before χ; from equation (5.5) χ cannot have a zero before ϕ as long as

$$1 - n\beta - m^2\beta^2 \geqslant 0. \qquad \ldots\ldots(5.8)$$

In the range $0 \leqslant r \leqslant r_0$, β has its maximum value at $r=r_0$, and for positive Y_0^2 it has its maximum value $2/m^2$ when $Y_0^2=0$. It can then be seen simply that inequality (5.8) is satisfied for all r in $(0, r_0)$ and all $Y_0^2>0$ if

$$m^2 \geqslant 2n+4. \qquad \ldots.(5.9)$$

Thus, if (5.9) is satisfied there are no solutions to the dispersion relation with positive Y_0^2, and no instabilities.

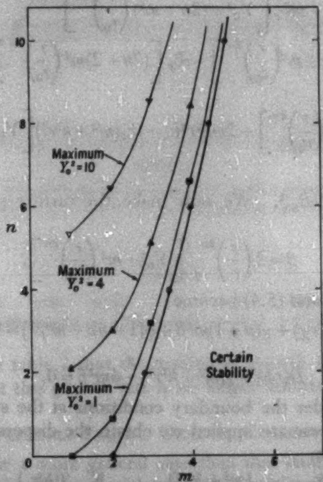


Figure 2. Plot in the m, n plane of the sufficiency condition for stability. The region of certain stability is marked, and other curves show the maximum possible growth rates in regions in which there may be instability.

In addition if $m^2 < 2n+4$ we can specify maximum possible instability growth rates Y_0^2. These are given by

$$1-\frac{2n}{m^2+Y_0^2}-\frac{4m^2}{(m^2+Y_0^2)^2}=0$$

or

$$Y_0^2=n-m^2+(n^2+4m^2)^{1/2}. \qquad \ldots\ldots(5.10)$$

When $n \gg m$ this leads to

$$Y_0^2 \simeq 2n-m^2. \qquad \ldots\ldots(5.11)$$

It should be stressed that the inequality (5.8) only provides sufficiency conditions for stability against a given perturbation. In particular, when it is not satisfied, we have neither proved that the configuration is unstable, nor that the maximum possible growth rate given by (5.10) or (5.11) is attained. However, we have seen, in the special case $m=1$, $n=0$ considered in the previous section, that the actual maximum growth rate is equal to the maximum possible predicted by equation (5.10). Figure 2 illustrates the sufficiency conditions discussed above.

§ 6. COMPRESSIBLE FLUID—GENERAL CURRENT DISTRIBUTION

In this section we extend the work of the previous section to the more realistic case of a compressible plasma in which the equilibrium density has the same distribution as the pressure. Thus, in equilibrium, the physical variables in the plasma have the values

$$\rho=\rho_0[1-(r/r_1)^{2n+2}],\ p=p_0[1-(r/r_1)^{2n+2}],$$

$$\mathbf{B}=[0,B_{\theta0}(r/r_0)^{n+1},0],\ \mathbf{j}=\left[0,0,\frac{(n+2)cB_{\theta0}}{4\pi r_0}\left(\frac{r}{r_0}\right)^n\right],$$

where

$$\frac{B_{\theta0}^2}{4\pi}\frac{n+2}{2n+2}\left(\frac{r_1}{r_0}\right)^{2n+2}=p_0,$$

and the fluid has ratio of specific heats γ.

The conducting fluid is surrounded by a medium of density $\rho_1=\rho_0[1-(r_0/r_1)^{2n+2}]$, pressure $p_1=p_0[1-(r_0/r_1)^{2n+2}]$, ratio of specific heats γ_1 and carrying magnetic field $B_\theta=B_{\theta0}r_0/r$.

Equations (2.2) and (2.10), applied in the plasma, combine to give

$$\tilde{\mathbf{B}}=\frac{iB_{\theta0}}{\omega r_0}(r/r_0)^n[m\tilde{v}_r,\ m\tilde{v}_\theta+ir\operatorname{div}\tilde{\mathbf{v}}+in\tilde{v}_r,\ m\tilde{v}_z]. \qquad \ldots\ldots(6.1)$$

Equations (2.7) and (2.8) combine to give one equation between $\tilde{p}$ and the components of $\tilde{\mathbf{v}}$

$$\tilde{p}=-\frac{ic_s^2\rho_0}{\omega}\left[1-\left(\frac{r}{r_1}\right)^{2n+2}\right]\operatorname{div}\tilde{\mathbf{v}}+\frac{2n+2}{\omega r_1}p_0\left(\frac{r}{r_1}\right)^{2n+1}\tilde{v}_r, \qquad \ldots\ldots(6.2)$$

while the second component of equation (2.5) leads to another such relation

$$\tilde{p}=\frac{i\rho_0\omega}{m}\left[1-\left(\frac{r}{r_1}\right)^{2n+2}\right]r\tilde{v}_\theta+\frac{(n+2)B_{\theta0}^2}{4\pi\omega r_0^2}\left(\frac{r}{r_0}\right)^{2n}r\tilde{v}_r. \qquad \ldots\ldots(6.3)$$

Comparing the two values of $\tilde{p}$ we obtain an expression for div $\tilde{\mathbf{v}}$ which can be substituted into equation (6.1). Thus

$$\tilde{\mathbf{B}}=\frac{iB_{\theta0}}{\omega r_0}\left(\frac{r}{r_0}\right)^n\left[m\tilde{v}_r,\ m\tilde{v}_\theta\left(1+\frac{\omega^2r^2}{m^2c_s^2}\right)+in\tilde{v}_r,\ m\tilde{v}_z\right].$$

Using equations (6.2) and (6.3) in conjunction with the other two components of equation (2.5) we obtain two equations for $\tilde{v}_r$ and $\tilde{v}_\theta$:

$$\begin{aligned} &D(r\tilde{v}_r)2m\left(\frac{r}{r_0}\right)^{2n} + m\tilde{v}_r\left[Y_0^2\left\{1-\left(\frac{r}{r_1}\right)^{2n+2}\right\} + (2n+m^2)\left(\frac{r}{r_0}\right)^{2n}\right] \\ &\quad + iD(r\tilde{v}_\theta)\left[Y_0^2\left\{1-\left(\frac{r}{r_1}\right)^{2n+2}\right\} + m^2\left(\frac{r}{r_0}\right)^{2n} + \frac{(2n+2)Y_0^2}{(n+2)\gamma}\left(\frac{r}{r_1}\right)^{2n+2}\right] \\ &\quad + i\tilde{v}_\theta\left[(2n+2)m^2\left(\frac{r}{r_0}\right)^{2n} + \frac{(4n+4)Y_0^2}{\gamma}\left(\frac{r}{r_1}\right)^{2n+2} - (2n+2)Y_0^2\left(\frac{r}{r_1}\right)^{2n+2}\right] = 0 \end{aligned}$$

and $$\qquad \ldots\ldots(6.4)$$

$$\begin{aligned} &mD(r\tilde{v}_r)\left[Y_0^2\left\{1-\left(\frac{r}{r_1}\right)^{2n+2}\right\} + m^2\left(\frac{r}{r_0}\right)^{2n}\right] + 2k^2r^2\left(\frac{r}{r_0}\right)^{2n} m\tilde{v}_r \\ &\quad + i\tilde{v}_\theta\left[\left\{m^2 + \frac{(2n+2)Y_0^2}{(n+2)\gamma}\left(\frac{r_0}{r_1}\right)^{2n}\left(\frac{r}{r_1}\right)^2\right\}\left\{Y_0^2\left[1-\left(\frac{r}{r_1}\right)^{2n+2}\right] + m^2\left(\frac{r}{r_0}\right)^{2n}\right\}\right. \\ &\quad \left. + k^2r^2\left[Y_0^2\left\{1-\left(\frac{r}{r_1}\right)^{2n+2}\right\} + m^2\left(\frac{r}{r_0}\right)^{2n} + \frac{(2n+2)Y_0^2}{(n+2)\gamma}\left(\frac{r}{r_1}\right)^{2n+2}\right]\right] = 0. \end{aligned}$$

$$\ldots\ldots(6.5)$$

We again introduce new variables $m\tilde{v}_r = \phi$, $\beta\phi + i\tilde{v}_\theta = \chi$, where now

$$\beta = 2\left(\frac{r}{r_0}\right)^{2n}\left[\left\{Y_0^2\left[1-\left(\frac{r}{r_1}\right)^{2n+2}\right] + m^2\left(\frac{r}{r_0}\right)^{2n} + \frac{(2n+2)Y_0^2}{(n+2)\gamma}\left(\frac{r}{r_1}\right)^{2n+2}\right\}\right]^{-1}.$$

We will not, however, rewrite equations (6.4), (6.5) in terms of ϕ and χ because of the very complicated form of some of the coefficients. Once again the dispersion relation for the problem may be quite simply expressed in terms of the values of ϕ and χ at $r = r_0$. In fact

$$\begin{aligned} \frac{\chi}{\phi} = &-\left[\left\{Y_0^2\left[1-\left(\frac{r_0}{r_1}\right)^{2n+2}\right] + m^2 + \frac{(2n+2)Y_0^2}{(n+2)\gamma}\left(\frac{r_0}{r_1}\right)^{2n+2}\right\}\right]^{-1} \\ &\times\left[\frac{\rho_1 Y_0^2}{\rho_0\zeta_1 r_0}\frac{\mathrm{K}_m(\zeta_1 r_0)}{\mathrm{K}_m'(\zeta_1 r_0)} + \frac{m^2}{|k|r_0}\frac{\mathrm{K}_m(|k|r_0)}{\mathrm{K}_m'(|k|r_0)}\right], \end{aligned} \qquad \ldots\ldots(6.6)$$

where $\zeta_1^2 = k^2 + \rho_1\omega^2/\gamma p_1$ as defined in § 2.

It can be readily verified that equations (5.3), (5.4) and (5.7) may be obtained from equations (6.4), (6.5) and (6.6) by letting all terms in r/r_1, r_0/r_1 tend to zero.

As before, the right-hand side of equation (6.6) is always positive for positive Y_0^2 and thus solutions of the dispersion relation with positive Y_0^2 can only occur when χ and ϕ have the same sign at $r = r_0$. Also near $r = 0$ the behaviour of ϕ and χ is the same as in the incompressible case, and ϕ and χ have opposite signs subject to the restrictions stated in § 5. Once again it is simple to see from equation (6.5) (expressed in terms of ϕ and χ) that, since the coefficient of χ is positive, ϕ cannot vanish before χ. From equation (6.4), χ cannot vanish before ϕ if the coefficient of ϕ is positive. The coefficient of ϕ is

$$f\left[Y_0^2\left\{1-\left(\frac{r}{r_1}\right)^{2n+2}\right\} + m^2\left(\frac{r}{r_0}\right)^{2n} + \frac{(2n+2)Y_0^2}{(n+2)\gamma}\left(\frac{r}{r_1}\right)^{2n+2}\right]^{-2},$$

where f can be reduced to the form

$$\begin{aligned} f = &Y_0^4\left[1-\left(\frac{r}{r_1}\right)^{2n+2}\right]\left[\left\{1-\left(\frac{r}{r_1}\right)^{2n+2}\right\} + \frac{2n+2}{(n+2)\gamma}\left(\frac{r}{r_1}\right)^{2n+2}\right] \\ &+ Y_0^2\left(\frac{r}{r_0}\right)^{2n}\left[1-\left(\frac{r}{r_1}\right)^{2n+2}\right](2m^2-2n) \\ &+ Y_0^2\left(\frac{r}{r_0}\right)^{2n}\left(\frac{r}{r_1}\right)^{2n+2}\frac{2n+2}{(n+2)\gamma}[m^2-2n-4] + m^2\left(\frac{r}{r_0}\right)^{4n}[m^2-2n-4]. \end{aligned}$$

It can be seen that f is positive for all r, $0 \leqslant r \leqslant r_0$ and for all $Y_0^2 \geqslant 0$, provided that

$$m^2 \geqslant 2n+4. \qquad \ldots\ldots(6.7)$$

Thus, exactly as in the case of an incompressible fluid of constant density, there can be no instability of azimuthal wave number m if $m^2 \geqslant 2n+4$.

In cases when the inequality (6.7) is not satisfied we can again specify maximum possible instability growth rates Y_0^2. However, in this case the limits we can set depend on how close r_0 is to r_1. If $r_0 = r_1$, then f is always negative at $r = r_1$ if $m^2 < 2n+4$. In this case no upper limit can be given by the arguments we are using, but as we are only obtaining sufficiency conditions we are also unable to state that the growth rate is unbounded. In general if $(r_0/r_1)^{2n+2} = 1-\lambda$, the maximum possible value of Y_0^2 will be proportional to $1/\lambda$. In the particular case $n=0$, $m=1$, considered in §4 the maximum value of Y_0^2 is given by the equation

$$Y_0^4\lambda\left[\lambda+\frac{1-\lambda}{\gamma}\right]+2\lambda Y_0^2-\frac{3}{\gamma}Y_0^2[1-\lambda]-3=0.$$

Without solving this exactly it is possible to see that the maximum possible value of Y_0^2 is less than $3/\lambda$.

In order to test whether the maximum possible growth rates obtained in the last two sections have any significance it would be necessary to solve the differential equations for ϕ and χ, and then to solve the dispersion relation. This would almost certainly require the use of an automatic computing machine. It is worth noting that the increased maximum growth rates obtained in this section are caused by removing the assumption of constant density and not directly by compressibility (though, of course, these two assumptions are bound up together).

§ 7. INFLUENCE OF VISCOSITY ON INSTABILITIES OF A SHEET CURRENT

In the study in §3 of the instabilities of an ideally conducting plasma bearing a sheet current, the most unsatisfactory feature is the infinite instability for large wavenumbers. In subsequent work on a distributed current the infinite growth rate is removed, but an instability at large wavenumbers remains. The introduction of viscosity will be expected to damp down such instabilities because of the large viscous stresses involved; the results of the following section for large wavenumber serve to confirm this view. In the case of very small wavenumbers no very large viscous forces will be called into play; hence we expect the relation between wavenumber and growth rate to remain unaltered for small wavenumber: this is also verified. As the instability growth rate is zero for both large and small wavenumbers, there will be a wavenumber of maximum instability. An approximate value of this wavenumber is obtained and we estimate its accuracy by using an analogy between this problem and that of hydrodynamic Rayleigh–Taylor instability (Chandrasekhar 1955, Hide 1955, Rayleigh 1883, Taylor 1950).

We consider an equilibrium configuration in which an ideally conducting plasma, of pressure p_0, density ρ_0, ratio of specific heats γ and coefficient of viscosity μ, bears a uniform cylindrical sheet current j_0^*. Surrounding this plasma is a vacuum. Because of our specification of this problem no magnetic field can penetrate into the plasma, and the coefficient of viscosity considered will be the normal coefficient of viscosity. In any actual arrangement, in which some volume current is carried by the plasma, there will be large magnetic fields, and the effect of the magnetic field on the transport coefficients will have to be considered.

We now consider perturbations of this equilibrium configuration of the particular form $\exp\{i(\theta+kz)+\omega t\}$. The equations in the plasma are

$$\rho\frac{d\mathbf{v}}{dt}=-\operatorname{grad}p+\tfrac{4}{3}\mu\operatorname{grad}\operatorname{div}\mathbf{v}-\mu\operatorname{curl}\operatorname{curl}\mathbf{v},$$

$$\frac{\partial\rho}{\partial t}=-\operatorname{div}\rho\mathbf{v},\qquad \frac{1}{p}\frac{dp}{dt}=\frac{\gamma}{\rho}\frac{d\rho}{dt}.$$

From these we can obtain a single equation for $\tilde{\mathbf{v}}$

$$\tilde{\mathbf{v}}=\left(\frac{c_s^2}{\omega^2}+\frac{4\nu}{3\omega}\right)\operatorname{grad}\operatorname{div}\tilde{\mathbf{v}}-\frac{\nu}{\omega}\operatorname{curl}\operatorname{curl}\tilde{\mathbf{v}}, \qquad \ldots\ldots(7.1)$$

where ν is the kinematic viscosity μ/ρ_0.
It can now be shown that $\tilde{\mathbf{v}}$ can be split into irrotational and solenoidal parts which independently satisfy equation (7.1).

For suppose $\tilde{\mathbf{v}}=\tilde{\mathbf{v}}_1+\tilde{\mathbf{v}}_2$ where $\tilde{\mathbf{v}}_1$ is irrotational and $\tilde{\mathbf{v}}_2$ is solenoidal. (We drop the superscript in the following discussion.)

Then
$$\left[\mathbf{v}_1-\left(\frac{c_s^2}{\omega^2}+\frac{4\nu}{3\omega}\right)\operatorname{grad}\operatorname{div}\mathbf{v}_1\right]=-\left[\mathbf{v}_2+\frac{\nu}{\omega}\operatorname{curl}\operatorname{curl}\mathbf{v}_2\right]. \qquad \ldots\ldots(7.2)$$

Taking the divergence of (7.2) we obtain

$$\operatorname{div}\left[\mathbf{v}_1-\left(\frac{c_s^2}{\omega^2}+\frac{4\nu}{3\omega}\right)\operatorname{grad}\operatorname{div}\mathbf{v}_1\right]=0, \qquad \ldots\ldots(7.3)$$

while
$$\operatorname{curl}\left[\mathbf{v}_1-\left(\frac{c_s^2}{\omega^2}+\frac{4\nu}{3\omega}\right)\operatorname{grad}\operatorname{div}\mathbf{v}_1\right]\equiv 0. \qquad \ldots\ldots(7.4)$$

Thus
$$\mathbf{v}_1-\left(\frac{c_s^2}{\omega^2}+\frac{4\nu}{3\omega}\right)\operatorname{grad}\operatorname{div}\mathbf{v}_1=\operatorname{grad}\psi \qquad \ldots\ldots(7.5)$$

and
$$\nabla^2\psi=0. \qquad \ldots\ldots(7.6)$$

It is now possible to write $\mathbf{v}=(\mathbf{v}_1-\operatorname{grad}\psi)+(\mathbf{v}_2+\operatorname{grad}\psi)$, such that $\mathbf{v}_1-\operatorname{grad}\psi$ and $\mathbf{v}_2+\operatorname{grad}\psi$ are both solutions of equation (7.1), while the first is irrotational and the second solenoidal.

The irrotational part of equation (7.1) gives acoustic waves only slightly modified by the presence of viscosity, and the solution can be expressed in terms of one constant $\tilde{p}_0$ where the perturbed pressure is $\tilde{p}=\tilde{p}_0\mathrm{I}_1(\zeta r)$
and

$$\zeta^2=k^2+\frac{\omega^2/c_s^2}{1+4\nu\omega/3c_s^2}.$$

The solenoidal velocity field gives rise to no density or pressure fluctuations and vanishes identically in the absence of viscosity. The solution is expressed in terms of two arbitrary constants $\tilde{v}_{r0}$ and $\tilde{v}_{z0}$, and Bessel functions of qr, where $q^2=k^2+\omega/\nu$. The complete solution for the perturbed plasma quantities is

$$\tilde{p}=\tilde{p}_0\mathrm{I}_1(\zeta r)\exp\{i(\theta+kz)+\omega t\},\quad \tilde{\rho}=\frac{\tilde{p}_0}{c_s^2}\mathrm{I}_1(\zeta r),$$

$$\tilde{v}_r=\frac{\tilde{v}_{r0}\mathrm{I}_1(qr)}{qr}-\frac{ik\tilde{v}_{z0}\mathrm{I}_2(qr)}{q}-A\zeta\tilde{p}_0\mathrm{I}_1{}'(\zeta r),$$

$$\tilde{v}_\theta=i\tilde{v}_{r0}\mathrm{I}_1{}'(qr)-\frac{k\tilde{v}_{z0}\mathrm{I}_2(qr)}{q}-\frac{iA\tilde{p}_0\mathrm{I}_1(\zeta r)}{r}$$

and
$$\tilde{v}_z=\tilde{v}_{z0}\mathrm{I}_1(qr)-ikA\tilde{p}_0\mathrm{I}_1(\zeta r),$$

where the exponential factor is understood in all quantities, ζ and q are both to have positive real parts and $A=\{1+4\nu\omega/3c_s^2\}/\rho_0\omega$.

When we apply the boundary conditions between the plasma and vacuum quantities, we can eliminate all the constants to obtain the dispersion relation. Because of its complexity we leave this relation in the form of three equations between $\tilde{p}_0$ $\tilde{v}_{r0}$ and $\tilde{v}_{z0}$:

$$\tilde{v}_{r0}[ik\mathrm{I}_1(qr_0)]+\tilde{v}_{z0}[k^2r_0\mathrm{I}_2(qr_0)+q^2r_0\mathrm{I}_1'(qr_0)]-2ikq\zeta r_0A\tilde{p}_0\mathrm{I}_1'(\zeta r_0)=0, \quad \ldots\ldots(7.7)$$

$$\begin{aligned}i\tilde{v}_{r0}[\mathrm{I}_1(qr_0)-qr_0\mathrm{I}_1'(qr_0)+q^2r_0^2\mathrm{I}_1''(qr_0)]+\tilde{v}_{z0}[2kr_0\mathrm{I}_2(qr_0)\\ -kqr_0^2\mathrm{I}_2'(qr_0)]+2iA\tilde{p}_0[q\mathrm{I}_1(\zeta r_0)-q\zeta r_0\mathrm{I}_1'(\zeta r_0)]=0\end{aligned} \quad \ldots\ldots(7.8)$$

and

$$\begin{aligned}&\frac{2p_0}{\omega r_0}\left\{1+\frac{1}{|k|r_0}\frac{\mathrm{K}_1(|k|r_0)}{\mathrm{K}_1'(|k|r_0)}\right\}\left\{\frac{\tilde{v}_{r0}\mathrm{I}_1(qr_0)}{qr_0}-\frac{ik\tilde{v}_{z0}\mathrm{I}_2(qr_0)}{q}-A\tilde{p}_0\zeta\mathrm{I}_1'(\zeta r_0)\right\}\\ &\quad=-\tilde{p}_0\mathrm{I}_1(\zeta r_0)+\frac{2\mu}{3}\left[\frac{\omega}{\rho_0c_S^2}\tilde{p}_0\mathrm{I}_1(\zeta r_0)\right]+2\mu\left\{\frac{\tilde{v}_{r0}\mathrm{I}_1'(qr_0)}{r_0}-\frac{\tilde{v}_{r0}\mathrm{I}_1(qr_0)}{qr_0^2}\right.\\ &\quad\left.-ik\tilde{v}_{z0}\mathrm{I}_2'(qr_0)-A\tilde{p}_0\zeta^2\mathrm{I}_1''(\zeta r_0)\right\}.\end{aligned} \quad \ldots\ldots(7.9)$$

In the absence of viscosity equations (7.7), (7.8) do not apply (strictly the left-hand side of these equations should contain a factor ν) and equation (7.9) reduces to

$$\frac{\rho_0r_0^2\omega^2|k|}{2p_0\zeta}\frac{\mathrm{I}_1(\zeta r_0)}{\mathrm{I}_1'(\zeta r_0)}=|k|r_0+\frac{\mathrm{K}_1(|k|r_0)}{\mathrm{K}_1'(|k|r_0)},$$

the equation used in § 3.

§ 8. VISCOSITY: SOLUTION OF DISPERSION RELATION

In the non-viscous treatment of this problem there is an infinite instability for infinite wavenumber. In fact we have seen that as $|k|\to\infty$

$$\omega^2\to\frac{2c_S^2}{\gamma r_0}|k|. \quad \ldots\ldots(8.1)$$

This result will be invalid in the presence of viscosity as such an instability would imply infinite stress components. An analogous problem to this one is the hydrodynamic Rayleigh–Taylor instability: this is the instability that can arise when a heavy fluid is resting on top of a light one. In the limiting case when the upper fluid is very much heavier than the lower and $|k|\to\infty$,

$$\omega^2\to g|k|. \quad \ldots\ldots(8.2)$$

Chandrasekhar (1955) and Hide (1955) have since considered this problem when the fluids are viscous and they find in the same limiting case as $|k|\to\infty$

$$\omega\to g/2\nu|k|, \quad \ldots\ldots(8.3)$$

so that the rate growth approaches zero as $|k|\to\infty$. It seems likely that a similar result holds in our problem and we will assume that as $|k|\to\infty$

$$\omega\to\delta/|k|, \quad \ldots\ldots(8.4)$$

where δ is a constant to be determined. Equation (8.4) is an assumption which must be justified.

We now introduce the non-dimensional variable $X_0=|k|r_0$, and using equation (8.4) we can write

$$qr_0=X_0+\frac{\alpha_1}{X_0^2}+\frac{\alpha_2}{X_0^3}+\mathrm{O}\left(\frac{1}{X_0^4}\right), \quad \ldots\ldots(8.5)$$

where $\alpha_1=\delta r_0^3/2\nu$.

Similarly

$$\zeta r_0 = X_0 + \frac{\beta_1}{X_0^3} + O\left(\frac{1}{X_0^4}\right).$$

Using equations (7.7) and (7.8) we now solve for $\tilde{v}_{r0}$ and $k\tilde{v}_{z0}/|k|$ in terms of $A\tilde{p}_0$, and a series of descending powers of X_0 to obtain

$$\tilde{v}_{r0} = \frac{X_0 A\tilde{p}_0}{r_0}\left[1 - \frac{\alpha_1}{X_0^2} + O\left(\frac{1}{X_0^3}\right)\right]$$

and

$$\frac{k\tilde{v}_{z0}}{|k|} = \frac{iX_0 A\tilde{p}_0}{r_0}\left[1 - \frac{\alpha_1}{X_0^2} + \frac{\beta_1 + \frac{1}{2}\alpha_1 - \alpha_2}{X_0^3} + O\left(\frac{1}{X_0^4}\right)\right].$$

Then we substitute these values into equation (7.9) and equate to zero the coefficient of the highest power of X_0 remaining. This leads to an equation for α_1 and hence for δ.
We obtain

$$\alpha_1 = c_s^2 r_0^2/2\gamma\nu^2 \qquad \ldots\ldots(8.6)$$

and

$$\omega = \frac{c_s^2}{\gamma\nu r_0|k|} + O\frac{1}{|k|^2}. \qquad \ldots\ldots(8.7)$$

It thus appears that our assumption of equation (8.4) is justified, and the growth rate does fall to zero as $|k| \to \infty$. Using equation (8.7) a further analogy with the case of Rayleigh–Taylor instability may be seen. Not only are equations (8.1) and (8.7) similar in form to equations (8.2) and (8.3), but the constants in the equations, g, $2c_s^2/\gamma r_0$, are changed in the same ratio by the introduction of viscosity.

We now consider the solution for small wavenumber. In this case no large viscous stresses are involved and we expect the solution asymptotically to equal the inviscid solution

$$\omega \to |k|(2Lc_s^2/\gamma)^{1/2}, \qquad \ldots\ldots(8.8)$$

with

$$L = \left(\ln\frac{2}{|k|r_0} - C + 1\right), \qquad \ldots\ldots.(8.9)$$

where C is Euler's constant.
We assume that ω has this form and verify that it satisfies equations (7.7) to (7.9). With this value of ω, $q \sim (\omega/\nu)^{1/2}$ and $\zeta \sim \omega/c_s$. Equations (7.7) and (7.8) then give solutions for $\tilde{v}_{r0}$ and $\tilde{v}_{z0}$ in terms of $A\tilde{p}_0$:

$$\tilde{v}_{r0} \sim A\tilde{p}_0 \nu\omega^2/c_s^2, \quad \tilde{v}_{z0} \sim A\tilde{p}_0 2i\nu^{1/2}\omega^{1/2}k/c_s.$$

When these results are substituted into equation (7.9) we find that all the terms peculiar to the viscous problem are of lower order than the non-viscous terms, and to first order the solution for ω is exactly that given by equation (8.8).

We have now obtained asymptotic solutions to the dispersion relation for large and small wavenumbers. Because of its complicated form we do not in this paper attempt any further solution of the equation. In a qualitative way we expect the solution to appear as in figure 3. The upper curve shows the solution in the absence of viscosity in the (λ, ω) plane where $\lambda = 2\pi/|k|$ is the wavelength of the disturbance. The asymptotic behaviour for large wavelength follows this curve. The straight line through the origin shows the asymptotic behaviour for small wavelength, and the qualitative behaviour of the entire solution may be expected to resemble the curve shown with M at the wavelength λ_M of maximum instability. What can be simply found is the point M′, also shown, which is the intersection of the curves of asymptotic behaviour at small wavelengths for the viscous and inviscid cases. The corresponding $\lambda_{M'}$ should be a first approximation to λ_M.

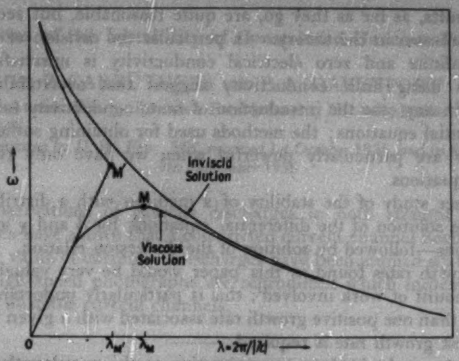


Figure 3. Qualitative plot of the relationship between growth rate and wavelength for the $m=1$ instability of a conducting plasma carying a surface current. This shows the inviscid solution and the effect of the introduction of viscosity. Also marked are the wavelength λ_M of maximum instability and the approximation to it, $\lambda_{M'}$, obtained in this paper.

To get some idea of the probable accuracy of this approximation we appeal once more to the analogy between our problem and that of Rayleigh–Taylor instability. In the case of Rayleigh–Taylor instability investigated by Chandrasekhar, the exact solution of the viscous dispersion relation can be obtained, and his λ_{MC} compared with $\lambda_{M'C}$. The result is that $\lambda_{MC}/\lambda_{M'C} \simeq 1{\cdot}3$, which means that the approximation is really a very good one, and leads us to believe that the approximate value should be of the right order of magnitude in this problem also.

Using the asymptotic forms for the inviscid and viscous problems we obtain

$$\lambda_{M'}/r_0 = 2\pi(2\gamma)^{1/3}(\nu/c_s r_0)^{2/3}.$$

If we now introduce from kinetic theory $\nu = \frac{1}{2}\bar{c}l$, where $\bar{c}$ is the mean molecular speed and l is a relevant mean free path, and also $c_s = 1{\cdot}086(\frac{1}{3}\gamma)^{1/2}\bar{c}$, we obtain $\nu/c_s r_0 = l/2{\cdot}172(\frac{1}{3}\gamma)^{1/2} r_0$.

Then $\lambda_{M'}/r_0 \simeq 6(l/r_0)^{2/3}$ for a monatomic gas with $\gamma = 5/3$.

§ 9. CONCLUSIONS

We are now in a position to see how the investigations of the later sections of the paper have modified the results of § 3. In the absence of viscosity the assumption of a distributed current leads to the following results:

(*a*) There are no highly twisted modes of instability unless n is large, on account of the stability criterion $m^2 > 2n + 4$.

(*b*) There is no infinite instability, as long as we realize it is unreasonable to allow the plasma density to be zero at its surface.

(*c*) Maximum possible growth rates can be obtained in terms of m, n and the ratio of surface to central plasma densities.

It should be realized that all conditions obtained are sufficiency conditions and we have not proved that maximum possible growth rates are attained. Using the results of § 4 it seems likely that the maximum growth rate still occurs at zero wavelength.

These results, as far as they go, are quite reasonable, but serious approximations still remain in the theory. In particular the division of the fluid into regions of infinite and zero electrical conductivity is unsatisfactory. First investigations using finite conductivity suggest that overstability might be important. In any case the introduction of finite conductivity raises the order of the differential equations; the methods used for obtaining sufficiency criteria in this paper are particularly powerful when we have only two first-order differential equations.

Any further study of the stability of a medium with a distributed current would require solution of the differential equations for ϕ and χ and tabulation of the functions—followed by solution of the dispersion relation. The limiting values for growth rates found in this paper would be very valuable in cutting down the amount of work involved; this is particularly important when there may be more than one positive growth rate associated with a given wavenumber, and the largest growth rate is required.

In § 8 we have obtained an approximation to the wavelength of maximum instability when viscosity has been introduced. This approximate value depends on a mean free path l which in the problem considered can be taken to be the ordinary kinetic theory mean free path, since there are no magnetic fields in the plasma. In this case the wavelength of maximum instability is probably of the same order of magnitude as the discharge radius (this, of course, depends on the precise relation between l and r_0). However, if magnetic fields in the plasma cause a large reduction in the relevant mean free path, then λ_M becomes small compared with the discharge radius. Once again the importance of finite conductivity and the effect of magnetic fields on transport coefficients needs to be investigated.

ACKNOWLEDGMENTS

I am grateful to several colleagues for many helpful discussions; I am especially indebted to Dr. W. B. Thompson who suggested the problem and to Miss S. J. Roberts who helped to check much tedious algebra.

REFERENCES

CHANDRASEKHAR, S., 1955, *Proc. Camb. Phil. Soc.*, **51**, 162.
DUNGEY, J. W., and LOUGHHEAD, R. E., 1954, *Aust. J. Phys.*, **7**, 5.
HIDE, R., 1955, *Proc. Camb. Phil. Soc.*, **51**, 179.
KRUSKAL, M., and SCHWARZSCHILD, M., 1954, *Proc. Roy. Soc.* A, **223**, 348.
LOUGHHEAD, R. E., 1955, *Aust. J. Phys.*, **8**, 319.
LORD RAYLEIGH, 1883, *Proc. Lond. Math. Soc.*, **14**, 170.
TAYLER, R. J., 1957, *Phil. Mag.*, **2**, in the press.
TAYLOR, G. I., 1950, *Proc. Roy. Soc.* A, **201**, 192.

Observations of the Instability of Constricted Gaseous Discharges

By R. CARRUTHERS and P. A. DAVENPORT
Atomic Energy Research Establishment, Harwell, Berks.

Communicated by D. W. Fry; *MS. received* 1st *October* 1956, *and in final form* 9*th November* 1956

Abstract. Observations of gaseous discharges in both straight and toroidal discharge tubes have shown that a constricted current channel becomes unstable at a particular value of current. Discharges have been studied up to currents of 1500 A and high-speed photographs are reproduced which indicate the helical form of these unstable current channels.

---

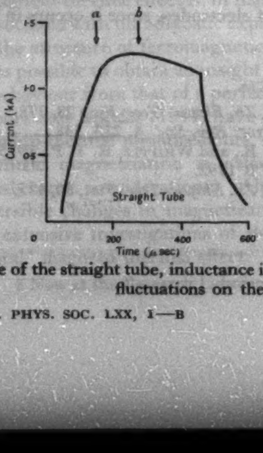
§ 1. Introduction

A straight flexible conductor carrying a current is unstable with respect to a local deformation. The average magnetic field over the cross section of the conductor at a small deformation no longer vanishes and a net body force acts to increase the deformation. The force per unit length is of the order of I^2/R, where I is the current and R the local radius of curvature. This phenomenon has been discussed theoretically by Kruskal and Schwarzschild (1954) and by Roberts and others (1955) and observed experimentally by Granovskii and Timofeeva (1955).

If the current passing through a column of gas at low pressure is increased, the radial distribution of current becomes increasingly peaked about the axis of the column, and a constricted axial current channel is formed. This is a manifestation of the well-known ' pinch effect '. We have observed that at sufficiently high currents (above 650 A in our apparatus) this current channel is unstable, and assumes a rapidly fluctuating helical configuration.

§ 2. Apparatus

Two types of discharge tube were used in these investigations. The first was a straight vertical cylindrical glass tube, 200 cm long and of 9·7 cm internal



In the case of the straight tube, inductance in the external circuit prevents the appearance of fluctuations on the current waveform.

diameter, with a water-cooled metal anode at the top and a mercury pool cathode at the bottom. An essentially rectangular voltage pulse of 500 μsec duration was applied between anode and cathode, the tube having been made conducting by means of a weak d.c. discharge to an auxiliary electrode. The mercury vapour pressure in the tube was about 10^{-3} mm Hg, and the mean longitudinal electric field during the pulse was about 0·4 v cm^{-1}. Pulse currents of 1200 A were obtained.

The second type of discharge tube consisted of a glass torus, 10 cm bore and of 30 cm mean diameter, which linked the core of a large pulse transformer. Gas in the torus was made conducting by means of a weak electrostatic radio-frequency discharge, and formed the single-turn secondary of the transformer. A rectangular voltage pulse of 500 μsec duration applied to the primary of the transformer produced an electric field of about 0·5 v cm^{-1} in the torus, and currents of 1500 A were obtained in argon and xenon at pressures of 1 to 5×10^{-3} mm Hg.

§ 3. OBSERVATIONS

The discharges were observed with an electronically gated image-converter tube which enabled photographs of 8 μsec exposure to be taken at any desired moment during the pulse. The accompanying photographs (Plate) show (*a*) the initial constriction and (*b*) the subsequent helical configuration of the current channel in both types of discharge. The moment of exposure is indicated on the current waveform in each case.

Photographs taken in two perpendicular directions show that the form of the current channel is helical and not sinusoidal. The mean of 27 photographic measurements of the wavelength of the helix in the straight tube was 19·1 cm; 18 photographs of the toroidal discharge all indicate six complete wavelengths round the torus. Both these results correspond to a pitch of approximately twice the internal diameter of the tube.

§ 4. DISCUSSION

Timofeeva and Granovskii (1956) have published photographs of transient arcs (duration 1 to 7μsec) in apparatus similar to, but smaller than, our straight tube. A helical current channel is discernible in some of these photographs. Our observations show that the helical configuration persists throughout long pulses of current which approach d.c. conditions and, further, that the phenomenon is not connected with the presence of electrodes, since it occurs in toroidal discharges.

REFERENCES

GRANOVSKII, V. L., and TIMOFEEVA, G. G., 1955, *Zh. Eksper. Teor. Fiz.*, **28**, 378.
KRUSKAL, M., and SCHWARZSCHILD, M., 1954, *Proc. Roy. Soc.* A, **223**, 348.
ROBERTS, S. J., STURROCK, P. A., THOMPSON, W. B., and WHIPPLE, R. T. P., 1955, Atomic Energy Research Establishment Report T/R 1792.
TIMOFEEVA, G. G., and GRANOVSKII, V. L., 1956, *Zh. Eksper. Teor. Fiz.*, **30**, 477.

Magneto-Thermal Effects in Nickel and Cobalt

By R. S. TEBBLE and R. W. TEALE†
Department of Physics, University of Leeds

Communicated by E. C. Stoner; *MS. received* 11*th June* 1956

Abstract. Comprehensive measurements on the reversible and total magneto-thermal effect, the reversible susceptibility and the magnetization curves of annealed nickel and cobalt have been made. It is shown that, for reversible changes in magnetization, there is no significant change in the value of the 'rotational' coefficient b'' for nickel in high fields, and that reversible changes in intrinsic magnetization and rotations of the magnetization vectors must predominate. For cobalt, with a higher coercivity, these effects also predominate even in relatively low fields. There is no evidence of the reduction in the numerical value of b'' in high fields, which has been deduced from the results of earlier work, and it is suggested that if any additional effects are present in high fields they must be irreversible. There is good agreement with the results of Tebble, Wood and Florentin and of Bates and his co-workers.

§ 1. INTRODUCTION

Although a certain amount of information about the nature of the change in magnetization of a ferromagnetic material may be gleaned from straightforward magnetic investigations, such as measurements of magnetization curves, reversible susceptibility and the Barkhausen effect, the application of the results of such work is limited. One of the most promising ways of obtaining a real understanding of the problem is from a study of the magneto-thermal effect—the change in temperature which accompanies an adiabatic change in the magnetization or, if the changes are reversible, $(\partial T/\partial H)_S$ in a standard notation. The value of $(\partial T/\partial H)_S$ depends on the physical nature of the change in magnetization, which may be brought about by processes such as changes in intrinsic magnetization, rotations of the magnetization vectors against the crystalline anisotropic forces ('rotations') and boundary movements. The magneto-thermal effect is, in many ways, analogous to the change in temperature associated with the adiabatic expansion (or compression) of a gas; information about the structure of ferromagnetic materials may be derived rather in the way that it is possible to obtain an insight into the structure of a gas from the deviation in its behaviour from that of a perfect gas.

Stoner and Rhodes (1949) have developed a method of analysing the results of magneto-thermal measurements, in which the effect of rotations and changes in intrinsic magnetization is calculated, using expressions developed from thermodynamic considerations. Although the treatment is in general applicable to reversible changes in magnetization only, it has been applied to the results of the extensive investigations of Bates and his co-workers (see Bates 1951) on the 'total' magneto-thermal effect. In these experiments the specimen is taken, step by step, through a magnetization curve and the temperature change is measured as the magnetization is changed adiabatically at each step. The temperature changes are small ($\sim 10^{-5}$ deg), and it is not possible to carry out these measurements completely under reversible conditions; in general both reversible and irreversible changes are produced. However, in high fields, where reversible effects predominate, the results are of considerable interest and it would appear that in some of the materials examined effects other than rotations and changes in intrinsic magnetization were present; this was indicated by a decrease in high fields in the numerical value of the coefficient b'' defined below. In low fields, in the hysteresis region, the changes in magnetization are largely irreversible and the method of analysis cannot be used; this has been appreciated, and methods have been devised for carrying out the magneto-thermal experiments under reversible conditions, in the work of Tebble, Wood and Florentin (1952) and Bates and Sherry (1955).

---

† Now at the Research Laboratories of Mullard Ltd., Salfords, Surrey.

With low coercivity materials, the measurements of the magneto-thermal effect under reversible conditions would involve the detection of temperature changes of the order 10^{-8} deg. This difficulty has been avoided in the method of Tebble, Wood and Florentin (1952) by measuring, not $(\partial T/\partial H)_S$, but the thermodynamically related quantity $(\partial I/\partial T)_H$ (see equation (7)). The results of this work on annealed nickel, coercivity $H_c = 0{\cdot}36$ oersted, indicated that in the field range 10 to 50 oersteds rotations and changes in intrinsic magnetization predominated, as shown by the constancy of b''. In the hysteresis region (below 10 oersteds) it was clear that other effects were present but it was not possible to come to any definite conclusions as to their character. An attempt was made to separate the reversible and irreversible components of the magneto-thermal effect by subtracting the integrated heat change, obtained from the reversible measurements, from the corresponding values given by Bates and Davis (1950) for the total heat change in a similar specimen. It was suggested that the presence of irreversible processes need not result in heating, as has sometimes been implied, and that the net cooling noticed in low fields, when the applied field is increased from zero, could be caused by irreversible changes in magnetization.

Bates and Sherry (1955) were able to separate the reversible and irreversible components of the change in magnetization in high coercivity materials (cobalt), by the direct measurement of the temperature changes produced under reversible conditions. This was possible because the coercivity of the cobalt was so high that, even with relatively large changes in field, the changes in magnetization were reversible, and measurable changes in temperature were produced. The limitations of this method are best described in a quotation from the conclusion in the paper of Bates and Sherry: "From the results on cobalt it is clear that the new direct method provides a satisfactory initial separation of the reversible and irreversible components. It is, however, limited to materials in which the incremental susceptibility is nowhere large; but even in less favourable cases it should be of use in giving a general indication of the components."

It became obvious during the course of the investigations of Tebble, Wood and Florentin (1952) (which will be referred to as I) that the scope of the work ought to be widened and that measurements on the reversible and total magneto-thermal effect, together with the relevant magnetic quantities, should be carried out on the same specimen. This has now been done, in the work described here, for annealed nickel and annealed cobalt, and the measurements on the reversible effect have been extended to high fields to enable the presence of effects other than rotations and changes in intrinsic magnetization in that region to be investigated.

§ 2. THE METHOD OF ANALYSIS

The change in the temperature T, produced by an adiabatic change in the field H and magnetization I, is usually given in the form of a heat change

$$Q' = \sum_{-H_{\max}}^{H} \delta Q' = \sum_{-H_{\max}}^{H} C\rho\, \delta T, \qquad \ldots\ldots(1)$$

where C and ρ are the specific heat and density of the material. It should be noted that Q' is the change in the thermal energy and not the heat intake $dQ = dE - HdI$, which in an adiabatic change is zero. We thus have

$$HdI = dE = dQ' + dE_M. \qquad \ldots\ldots(2)$$

When the magnetization is changed adiabatically, the work done, HdI, is equal to the change in internal energy, which is made up of a change in the 'magnetic' energy and a change in the 'thermal' energy (i.e. a change in temperature).

In the Stoner and Rhodes method of analysis, the measured heat change is written

$$Q' = a\int d(IH) + \int b'' H dI \qquad \ldots\ldots(3)$$

where the first term represents that part of the heat change associated with the change in intrinsic magnetization and

$$a = -\frac{T}{I_0}\frac{dI_0}{dT} \qquad \ldots\ldots(4)$$

is normally positive in sign; the value of a can be evaluated with reasonable accuracy. If the only other processes present are reversible rotations in the anisotropic field of the crystal, then the second term may be written $b\int HdI$, with

$$b = \frac{T}{K}\frac{dK}{dT} \qquad \ldots\ldots(5)$$

and K the modulus of the anisotropy coefficient; b is usually negative. The two effects represented by the two terms in equation (3) are thus of opposite sign, with changes in intrinsic magnetization predominating in high fields and the effect of rotations increasing in importance at low fields. From a comparison of the value of b given by equation (5) with the value of b'' given by equation (3) i.e.

$$b'' = \frac{(\partial Q'/\partial H)_S - a\{I + H(\partial I/\partial H)_T\}}{H(\partial I/\partial H)_T} = \frac{(\partial Q'/\partial H)_S - a(I + H\kappa_{\mathrm{rev}})}{H\kappa_{\mathrm{rev}}} \qquad \ldots\ldots(6)$$

(κ_{rev} the reversible susceptibility), it should be possible to detect the presence of effects other than rotations and changes in the intrinsic magnetization. With no other process than these taking place, b'' should be equal to b and be sensibly constant over the reversible part of the magnetization curve.

In the present work the reversible magneto-thermal coefficient $(\partial Q'/\partial H)_S$ is obtained from measurements on the reversible change in magnetization with temperature, using the relation

$$(\partial Q'/\partial H)_S = -T(\partial I/\partial T)_H \qquad \ldots\ldots(7)$$

(see Stoner and Rhodes 1949, p. 483). The method of measuring the total magneto-thermal effect is basically the same as that of Bates and his co-workers, i.e. the temperature change produced by an adiabatic change in field is measured directly by means of thermocouples; the arrangement of the thermocouples and the method of measuring the thermal e.m.f. is, however, different. The magnetization and reversible susceptibility are also required in the method of analysis, and these quantities are obtained using the standard methods described below.

In order to differentiate between the 'total' and reversible magneto-thermal coefficients, these will be written $(\partial Q'/\partial H)_{\mathrm{tot}}$ and $(\partial Q'/\partial H)_{\mathrm{rev}}$ respectively.

§ 3. EXPERIMENTAL METHOD

The apparatus has been designed so that it is possible to carry out measurements on the magnetization curve, reversible susceptibility, $(\partial Q'/\partial H)_{\mathrm{tot}}$ and $(\partial Q'/\partial H)_{\mathrm{rev}}$, in that order, without disturbing the specimen, thus reducing the possibility of any alteration in the magnetic properties of the material during the course of the experiments. The experimental arrangement is shown in figure 1, with the ellipsoidal specimen, 40 cm × 4 mm, mounted vertically in a Dewar flask, on which is wound the primary coil used in the reversible susceptibility measurements; the secondary coil is wound on the specimen. The thermocouple used in the $(\partial I/\partial T)_H$ measurement, and the 14 thermocouples used in the total magneto-thermal measurements, are attached to the specimen before it is inserted into the flask. The magnetizing coil is wound on a copper tube, into which the Dewar flask fits, and cooling is provided as shown in figure 1.

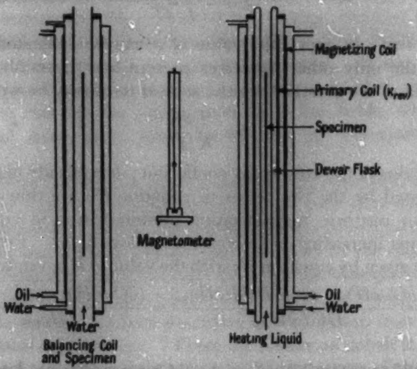


Figure 1. Experimental arrangement; the search coil and thermocouple are not shown.

3.1. *Reversible Magneto-Thermal Measurements*; $(\partial Q'/\partial H)_{\mathrm{rev}} = -T(\partial I/\partial T)_H$

The change in magnetization produced by a change in temperature in the specimen was measured by means of an astatic magnetometer placed as shown in figure 1. The magnetometer consists of an astatic pair of magnets separated by a distance of 20 cm; with this separation, each magnet is near the position at which the horizontal component of the field due to the ellipsoid is a maximum and each is subject to a couple of the same sense. The sensitivity of the system can be checked with the aid of a small calibrating coil mounted on the magnetometer. The temperature of the specimen was raised or lowered by passing heated, or cooled, liquid through the Dewar flask and the temperature change measured with a thermocouple attached to the specimen and a galvanometer.

The change in magnetization corresponding to a given change in field at the magnetometer can be calculated, knowing the sensitivity of the magnetometer and the dimensions and position of the specimen (see Appendix). Unfortunately, the total error arising from the considerable number of measurements which are required is large, and an alternative method of calibration has been devised. Using the search coil and galvanometer circuit used in obtaining the magnetization curves, the change in magnetization as the specimen is taken from a to b and b to c in figure 2 is measured. The magnetometer deflection corresponding to ac is also noted so that, with the magnetic field zero at the beginning and end of the change, the magnetometer can be calibrated directly in terms of change in the magnetic moment of the specimen; a deflection of 1 mm at 1 metre corresponded to a change in I in the specimen of about 10^{-3} e.m.u. during most of the measurements.

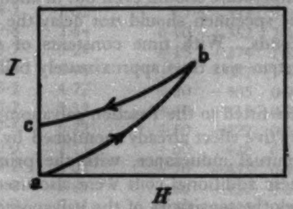


Figure 2. Changes in magnetization during the calibration.

3.2. *The Total Magneto-Thermal Measurements*; $(\partial Q'/\partial H)_{\mathrm{tot}}$

There are two main difficulties in this part of the experiment: firstly, the smallness of the temperature changes to be measured ($\sim 10^{-5}$ deg) and, secondly, the relatively large e.m.f. induced in the thermocouple leads by the change in induction in the solenoid and the specimen.

The e.m.f. produced in a single copper–constantan couple by a temperature change of 10^{-5} deg is about $4 \times 10^{-4}\,\mu$v. The method of using a number of thermocouples in series is complicated by the fact that the 'hot' junctions must be in good thermal contact with the specimen, which is itself an electrical conductor. Bates attached his junctions directly to the specimen, taking each pair of leads to a separate primary winding on a transformer with the secondary winding connected to a galvanometer. This method has the disadvantage that each pair of leads picks up induced e.m.f.s. In the present work the thermocouples are connected in series but the 'hot' junctions are insulated from the specimen by means of a thin layer of varnish which, it is calculated, does not materially affect the measurements; only one pair of leads is thus required. The temperature of the specimen was maintained relatively constant by the combined effect of the insulation of the Dewar flask and the cooling provided for the coils.

The specimen was coated with a thin layer of Araldite varnish, about 1/70 mm thick. The 14 thermocouples were made of strips of copper, 2 cm long, and constantan, 1 cm long, rolled flat from 28 s.w.g. wire into lengths about 1 mm wide by 1/10 mm thick. The thermal time constant of the copper strips (i.e. the time taken for the transmission of heat from one end to the other) is about 2 seconds, and that of the constantan about 10 seconds (Carslaw and Jaeger 1947). Ideally, these times should be equal, but there are limits to the dimensions of the thermocouples set by the method of construction. A blob of soft solder, about 2 mm in diameter, was attached to each cold junction to help stabilize its temperature.

The e.m.f. generated by the thermocouples was measured with a Paschen moving magnet galvanometer, with a resistance of 0·75 Ω matching, approximately, the total resistance of the thermocouple ~1 Ω. With the galvanometer shielded in a mumetal box, sensitivities of 4×10^3 mm $(\mu\mathrm{v})^{-1}$ were obtainable with a galvanometer period of 5 seconds.

It is of some importance that the order of the times associated with the effects involved in the detection of the temperature changes should be known. A change in induction in the thickest part of the specimen takes place in about 20 milliseconds, with susceptibilities of the order 100, and any variations in the resultant change in temperature should even out in about 70 milliseconds. The layer of varnish on the specimen should not delay the transmission of heat by more than 5 milliseconds. With time constants of several seconds for the thermocouples, the system was thus approximately ballistic; the galvanometer was critically damped.

Two extra coils were fitted to the Paschen galvanometer and these were used to balance out the inductive effect already mentioned by connecting them to the secondary coil of a mutual inductance, with the primary in series with the magnetizing coil. These additional coils were also used to adjust the galvanometer zero and to check the sensitivity of the galvanometer, without interfering with the thermocouple circuit which was extremely sensitive to extraneous effects. The apparatus was calibrated using the relation $(\partial Q'/\partial H)_S = -T(\partial I/\partial T)_H$. It was assumed that in high fields the changes in magnetization are reversible and the total magneto-thermal change Q'_{tot} is given by

$$Q'_{\mathrm{tot}} = \int_{H_1}^{H_2} (\partial Q'/\partial H)_{\mathrm{tot}}\, dH = -\int_{H_1}^{H_2} T(\partial I/\partial T)_H\, dH \qquad \ldots\ldots(8)$$

with the value of $(\partial I/\partial T)_H$ taken from the reversible magneto-thermal measurements. The sensitivity of the galvanometer was set so that a deflection of 1 mm at 1 metre corresponded to a thermal change of 270 erg cm^{-3} or 7×10^{-6} deg for nickel and 440 erg cm^{-3} or 11×10^{-6} deg for cobalt.

3.3. *Magnetization Curves*

The magnetization curves were obtained by the standard method using a Tinsley long period galvanometer as a fluxmeter; a mutual inductometer was used to calibrate the apparatus.

3.4. *Reversible Susceptibility* κ_{rev}

In high fields, where the changes in magnetization are mainly reversible, κ_{rev} was taken as the gradient of the (I, H) curve. In low fields it is more convenient to measure the reversible susceptibility, using the a.c. bridge method described by Tebble and Corner (1950).

§ 4. RESULTS

Measurements were made on specimens of spectroscopically pure nickel, annealed for 5 hours at 900°C, and cobalt, annealed for 9 hours at 900°C. Before annealing each was turned into the form of an ellipsoid 40 cm long and of 4 mm maximum diameter, giving a demagnetizing factor of 0·00540 (Stoner 1945); the coercivity of the nickel was 2·1 oersteds and of the cobalt 28 oersteds. The results are tabulated for convenience in any further calculations in tables 1 and 2; the values given have been taken from the best curves drawn through the experimental points. The results of the reversible magneto-thermal measurements in *I* on a specimen of nickel with a coercivity of 0·36 oersted are shown with the present results in figure 3, for the purpose of comparison.

Table 1. Nickel Specimen. Numerical Values of Measured Quantities.
(*a*) Low Field Region. The units for Q' are erg cm^{-3}.

H	I	κ_{rev}	$\left(\frac{\partial Q'}{\partial H}\right)_{\text{rev}}$	$Q'_{\text{tot}} \times 10^{-3}$	H	I	κ_{rev}	$\left(\frac{\partial Q'}{\partial H}\right)_{\text{rev}}$	$Q'_{\text{tot}} \times 10^{-3}$
−150	−454	0·155	+36	0	+2	−10	8·95	−27·0	15·02
−140	−452	0·170	+43·6	0·43	+3	+117	7·84	−24·0	14·37
−130	−450	0·191	+51·1	0·91	+4	+242	5·67	−23·0	14·02
−120	−448	0·219	+58·4	1·44	+5	+290	4·34	−23·0	13·72
−110	−446	0·256	+64·2	2·02	+7·5	+333	3·02	−25·5	13·28
−100	−443	0·293	+69·8	2·64	+10	+352	2·44	−29·4	12·82
−90	−440	0·335	+73·7	3·29	+15	+374	1·81	−36·8	12·07
−80	−437	0·375	+75·4	3·98	+20	+387	1·45	−43·7	11·53
−70	−433	0·444	+75·2	4·72	+30	+405	0·945	−56·8	10·64
−60	−428	0·523	+75·2	5·52	+40	+415	0·766	−67·6	9·84
−50	−422	0·627	+67·5	6·32	+50	+422	0·627	−76·0	9·05
−40	−416	0·766	+60·0	7·13	+60	+428	0·523	−79·5	8·29
−30	−409	0·945	+50·2	8·02	+70	+433	0·444	−79·7	7·55
−20	−395	1·45	+34·0	8·89	+80	+437	0·375	−76·9	6·81
−15	−386	1·81	+24·0	9·33	+90	+440	0·335	−72·7	6·12
−10	−371	2·36	+11·0	9·98	+100	+443	0·293	−67·6	5·46
−7·5	−362	2·75	+5·0	10·35	+110	+446	0·256	−62·1	4·82
−5·0	−348	3·35	−2·0	10·74	+120	+448	0·219	−56·0	4·22
−2·5	−328	4·33	−12·0	11·33	+130	+450	0·191	−49·6	3·68
0	−280	6·48	−21·0	12·42	+140	+452	0·170	−42·8	3·14
+1	−200	8·17	−25·0	13·22	+150	+454	0·155	−35·5	2·67

(*b*) High Field Region

H	I	κ_{rev}	$\left(\frac{\partial Q'}{\partial H}\right)_{\text{rev}}$	H	I	κ_{rev}	$\left(\frac{\partial Q'}{\partial H}\right)_{\text{rev}}$
+500	+468	0·008		+320	+466·5	0·034	+43
+480	+468	0·009	+60	+300	+466	0·041	+38
+460	+468	0·010	+60	+280	+465·5	0·050	+33
+440	+468	0·012	+60	+260	+464·5	0·060	+27
+420	+467·5	0·013	+58	+240	+463·5	0·076	+21
+400	+467·5	0·015	+55	+220	+462·5	0·089	+13
+380	+467·5	0·018	+53	+200	+460·5	0·108	+3
+360	+467	0·022	+50	+180	+458·5	0·135	−10
+340	+467	0·028	+47	+160	+456	0·178	−24

The calibration of the total magneto-thermal apparatus was carried out using equation (8); it was assumed that, for the annealed nickel, the total heat change

Q'_{tot} between 150 and 70 oersteds was equal to

$$\int_{150}^{70} \left(\frac{dQ'}{dH}\right)_{\text{rev}} dH$$

the corresponding field range for cobalt was 400 to 300 oersteds. An independent check on the accuracy of this method may be made by the application of Warburg's law i.e. the total heat evolved in one half cycle is equal to the hysteresis loss $\frac{1}{2}\oint H dI$. For nickel the value of Q'_{tot} at 150 oersteds is 2670 erg cm^{-3} compared with the value given by Warburg's law, 1800 erg cm^{-3}, a difference of 870 erg cm^{-3}; with the sum $\Sigma\,|\Delta Q'_{\text{tot}}|$ equal to 27 650 erg cm^{-3}, this represents a difference of 3%. The corresponding figures for cobalt were: at 500 oersteds $Q_{\text{tot}} = 24\,000$ erg cm^{-3}, $\frac{1}{2}\oint H dI = 20\,400$, giving a difference of 3600 ergs in 134 000, i.e. 3%.

Table 2. Cobalt Specimen. Numerical Values of Measured Quantities. The units for Q' are erg cm^{-3}.

H	I	κ_{rev}	$\left(\frac{\partial Q'}{\partial H}\right)_{\text{rev}}$	$Q'_{\text{tot}} \times 10^{-4}$
−500	−542			0
−480	−533			0·28
−460	−523			0·63
−440	−515	0·386		0·98
−420	−508	0·393		1·33
−400	−498	0·401	+181	1·68
−380	−490	0·411	+175	2·02
−360	−479	0·421	+169	2·36
−340	−469	0·432	+163	2·70
−320	−458	0·445	+157	3·02
−300	−448	0·458	+150	3·35
−280	−436	0·472	+143	3·68
−260	−424	0·489	+137	3·99
−240	−411	0·505	+130	4·27
−220	−397	0·522	+122	4·55
−200	−383	0·545	+114	4·82
−180	−370	0·569	+106	5·09
−160	−354	0·598	+97	5·34
−140	−339	0·623	+89	5·57
−120	−322	0·654	+79	5·80
−100	−303	0·691	+69	6·01
−80	−281	0·734	+58	6·21
−60	−257	0·787	+47	6·39
−40	−228	0·851	+34	6·58
−20	−193	0·935	+20	6·75
−10	−171	0·988	+11	6·83
0	−148	1·050	0	6·92
+5	−133	1·081		
+10	−119	1·122	−15	7·02
+15	−98	1·170		
+20	−67	1·225	−27	7·12
+25	−30	1·230		
+30	+20	1·160	−36	7·23
+40	+93	1·060	−44	7·32
+60	+165	0·925	−58	7·55
+80	+217	0·839	−70	7·79
+100	+251	0·773	−80	7·90
+120	+279	0·722	−89	7·90
+140	+301	0·677	−98	7·81
+160	+322	0·638	−107	7·66
+180	+340	0·601	−115	7·48
+200	+360	0·571	−123	7·28
+220	+373	0·544	−129	7·09
+240	+389	0·520	−137	6·87
+260	+403	0·498	−144	6·64
+280	+419	0·480	−150	6·39
+300	+432	0·462	−156	6·13
+320	+446	0·450	−161	5·88
+340	+458	0·437	−167	5·56
+360	+470	0·423	−172	5·25
+380	+481	0·411	−178	4·90
+400	+493	0·400	−183	4·53
+420	+504	0·391		4·14
+440	+513	0·382		3·74
+460	+522			3·31
+480	+532			2·86
+500	+542			2·40

The results of Bates and Davis (1950) on the total magneto-thermal effect in a specimen of annealed nickel, coercivity 2·0 oersteds, and of Bates and Sherry (1955) on annealed cobalt, coercivity 34 oersteds, are also shown in the appropriate figures, for comparison with the present measurements.

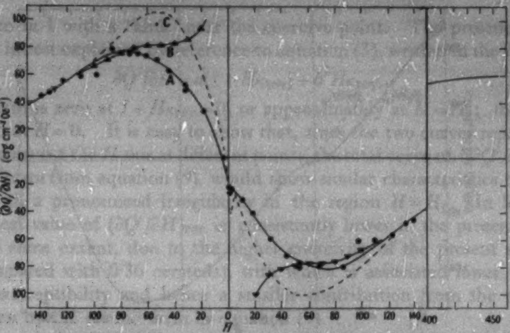


Figure 3. Magneto-thermal effect in annealed nickel ($H_c = 2{\cdot}1$). A, $(\partial Q'/\partial H)_{\text{rev}} = -T(\partial I/\partial T)_H$. $H < 150$ Oe, $\theta = 20{\cdot}5$°c; $H > 400$ Oe, $\theta = 28$°c. B, $(\partial Q'/\partial H)_{\text{tot}}$ for the same specimen. C, $(\partial Q'/\partial H)_{\text{rev}}$ for another specimen ($H_c = 0{\cdot}36$) taken from I.

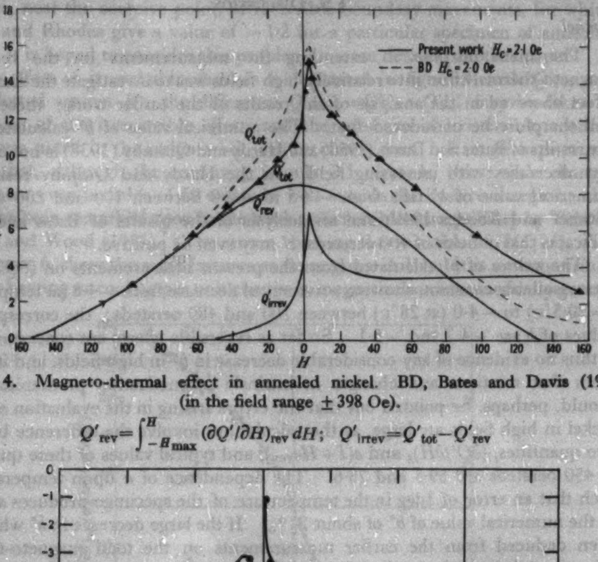


Figure 4. Magneto-thermal effect in annealed nickel. BD, Bates and Davis (1950) (in the field range ±398 Oe).

$$Q'_{\text{rev}} = \int_{-H_{\max}}^{H} (\partial Q'/\partial H)_{\text{rev}}\, dH; \quad Q'_{\text{irrev}} = Q'_{\text{tot}} - Q'_{\text{rev}}$$

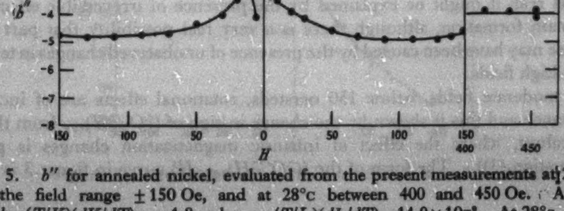


Figure 5. b'' for annealed nickel, evaluated from the present measurements at 20·5°c in the field range ±150 Oe, and at 28°c between 400 and 450 Oe. At 20·5°c $b = (T/K)(dK/dT) = -4{\cdot}8$ and $a = -(T/I_0)(dI_0/dT) = 14{\cdot}9 \times 10^{-2}$. At 28°c, $b = -5{\cdot}2$ and $a = 16{\cdot}2 \times 10^{-2}$.

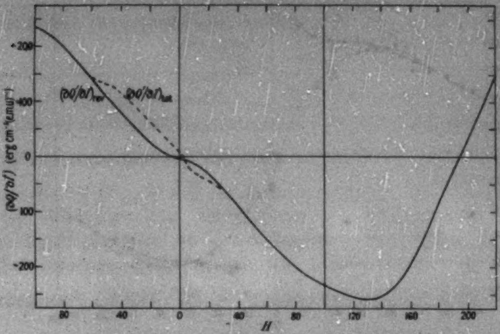


Figure 6. Variation of $(\partial Q'/\partial I)_{\text{rev}}$ and $(\partial Q'/\partial I)_{\text{tot}}$ with field for annealed nickel. The rate of change of internal magnetic energy with magnetization is $(\partial E_M/\partial I)_S = H - (\partial Q'/\partial I)_S$.

§ 5. DISCUSSION

Nickel.

The main purpose in extending the measurements on the reversible magneto-thermal effect into relatively high fields was to investigate the anomalous effect observed in the analysis of the results of the earlier work; these results will therefore be considered first. The numerical value of b'' calculated from the results of Bates and Davis (1950) and Hardy and Quimby (1938) is not constant but decreases with increasing field. In the Hardy and Quimby results the numerical value of b'' falls from $-4\cdot5$ to $-2\cdot0$ between 100 and 200 oersteds (Stoner and Rhodes 1949) and an analysis of the results of Bates and Davis indicates that in fields of 400 oersteds, b'' may even be positive.

The values of b'' calculated from the present measurements on $(\partial Q'/\partial H)_{\text{rev}}$ are reasonably constant, showing a numerical decrease from $-4\cdot8$ (at temperature $\theta = 20\cdot5$°c) to $-4\cdot0$ (at 28°c) between 100 and 480 oersteds; the corresponding values of b are $-4\cdot7$ and $-5\cdot1$. So far as reversible effects are concerned there is thus no evidence of any considerable decrease in b'' in high fields, and it would seem that rotations and changes in intrinsic magnetization predominate. It should, perhaps, be pointed out that the errors arising in the evaluation of b'' for nickel in high fields are large, as the calculations involve the difference between two quantities, $(\partial Q'/\partial H)_S$ and $a(I + H\kappa_{\text{rev}})$, and typical values of these quantities at 450 oersteds are 59·3 and 78·6. The dependence of a upon temperature is such that an error of 1deg in the temperature of the specimen produces an error in the numerical value of b'' of about $2\frac{1}{2}$%. If the large decrease in b'' which has been deduced from the earlier measurements on the total magneto-thermal effect is real, it might be explained by the presence of irreversible effects such as domain formation, although there is a very real possibility that part of this decrease may have been caused by the presence of unobserved changes in temperature in high fields.

In moderate fields, below 150 oersteds, rotational effects are of increasing importance and this is shown by the change in sign of $(\partial Q'/\partial H)_{\text{rev}}$ from the high field values, where the effect of intrinsic magnetization changes is greatest (cf. equation (3)). The form of the $((\partial Q'/\partial H)_{\text{rev}}, H)$ curve in figure 3 is similar to that given in I with a 'kink' near the coercive point. The presence of this irregularity is best explained by reference to equation (3), written in the form

$$\partial Q'/\partial H = a(I + H\kappa_{\text{rev}}) + b''H\kappa_{\text{rev}}. \qquad \ldots\ldots(9)$$

The first term is zero at $I + H\kappa_{\text{rev}} = 0$, or approximately at $H = H_c$; the second term is zero at $H = 0$. It is easy to show that, since the two curves representing these two terms cut the H axis at different points, the total curve of $((\partial Q'/\partial H)_s, H)$ for nickel, drawn from equation (9), would show similar characteristics to that in figure 3, with a pronounced irregularity in the region $H = H_c$. In low fields the numerical value of $(\partial Q'/\partial H)_{\text{rev}}$ is consistently lower in the present results. This is, to some extent, due to the higher coercivity of the present specimen (2.1 as compared with 0·36 oersteds), with which is associated lower values of reversible susceptibility and hence a smaller contribution from the rotational effects, given by the second term in equation (9).

The numerical value of b'' remains reasonably constant down to 60 oersteds then decreases fairly rapidly to a value of $-3{\cdot}2$ at 10 oersteds. It has been suggested by Tebble, Wood and Florentin (I) that this effect is caused by the presence, near the coercive point, of reversible boundary movements, for which Stoner and Rhodes give a value of $-1{\cdot}2$ for a particular specimen of annealed nickel; it is hoped to consider this problem in more detail in another paper.

The effect of irreversible changes is illustrated in figure 4, with Q'_{rev} and Q'_{tot} plotted against H; the difference $Q'_{\text{irrev}} = Q'_{\text{tot}} - Q'_{\text{rev}}$ may be taken as a measure of the irreversible effect, and the presence of irreversible cooling, which was noted in the earlier work, can be clearly seen. It is perhaps worth noting that, although the coercivity is only 2·1 oersteds, irreversible changes are present even in fields of 50 oersteds, well outside the hysteresis region; this effect has been remarked upon in the work on reversible susceptibility by Tebble, Corner and Wood (1951). The curves showing the variation of $(\partial Q'/\partial I)_s$ with field, figure 6, show the same features as those given in I, but it is now confirmed that the heating or cooling per unit change in magnetization is numerically greater

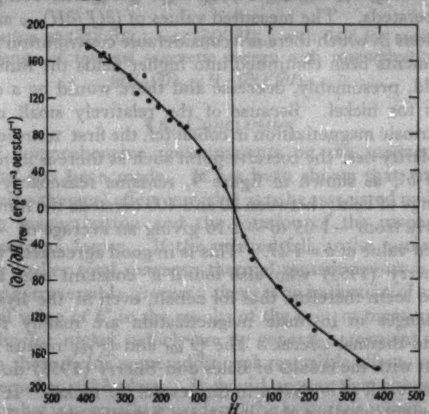


Figure 7. Reversible magneto-thermal effect in annealed cobalt.

for irreversible than for reversible changes, although there is no difference in order of magnitude; in the earlier work it was suspected that the difference might be due to experimental errors. The rate of change of internal magnetic energy with field is given by $(\partial E_M/\partial I) = H - (\partial Q'/\partial I)_s$ and, from figure 6, is positive for positive values of magnetic field, for both reversible and irreversible processes.

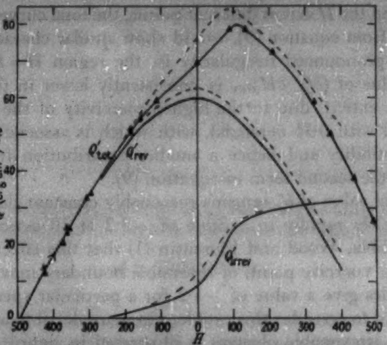


Figure 8. Magneto-thermal effect in annealed cobalt. The full curves are from the results of the present measurements; $H_c = 28$ Oe. The broken curves are taken from the measurements of Bates and Sherry (1955) in the field range ± 380 Oe; $H_c = 34$ Oe.

$$Q'_{\text{rev}} = \int_{-H_{\max}}^{H} (\partial Q'/\partial H)_{\text{rev}}\, dH, \quad Q'_{\text{irrev}} = Q'_{\text{tot}} - Q'_{\text{rev}}.$$

Cobalt.

The results for the cobalt specimen must necessarily be of more limited significance than those for nickel, because the maximum applied field of 500 oersteds can only be described as low or moderate, compared with the coercivity of 28 oersteds. The measured values of $(\partial Q'/\partial H)_{\text{rev}}$ are thus appropriate to those regions in which there is a considerable contribution from rotations: had the measurements been continued into higher fields the numerical value of $(\partial Q'/\partial H)_{\text{rev}}$ would, presumably, decrease and there would be a change in sign, as in the results for nickel. Because of the relatively small contribution of changes in the intrinsic magnetization in cobalt (cf. the first term in equation (9)), there is no irregularity near the coercive point such as there is in nickel.

The value of b'', as shown in figure 9, remains reasonably constant over a considerable range of field; between 20 and 400 oersteds the numerical value of b'' increases slowly from $-1{\cdot}05$ to $-1{\cdot}16$ giving an average of $-1{\cdot}1$ compared with the calculated value of $b = 1{\cdot}27$. This is in good agreement with the results of Bates and Sherry (1955) who also found b'' constant over a similar range of field. It would seem therefore that for cobalt, even in the hysteresis region, rotations and changes in intrinsic magnetization are mainly responsible for reversible magneto-thermal effects. The Q'_{tot} and Q'_{rev} results which are also in good agreement with the results of Bates and Sherry (1955) do not show any irreversible cooling such as is shown in the results for nickel. It is not possible to say at the moment whether this difference between the two materials is of significance as there is not, at present, any complete picture of the energy changes associated with the irreversible processes; in this instance the processes involved are probably irreversible boundary movements. The difference between the thermal effect produced by reversible and irreversible changes in magnetization is shown, in the $(\partial Q'/\partial I)_s$ curves in figure 10, to be extended over a wide range of field, in keeping with the relatively high coercivity.

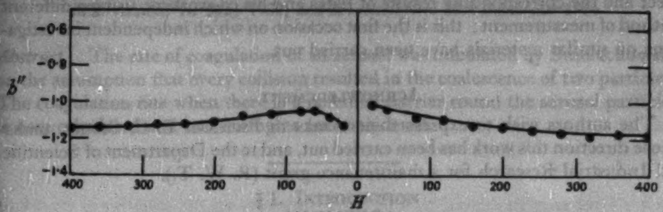


Figure 9. b'' for annealed cobalt, evaluated from the present measurements at 19°c; at 17°c $b=(T/K)(dK/dT)=-1{\cdot}2_7$, and $a=-(T/I_0)(dI_0/dT)=0{\cdot}8_8\times10^{-2}$.

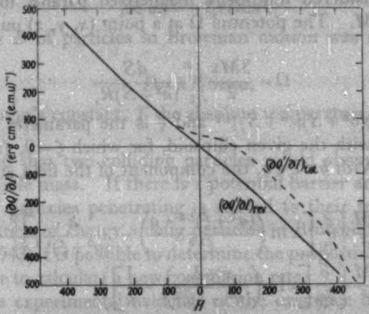


Figure 10. Variation of $(\partial Q'/\partial I)_{\text{rev}}$ and $(\partial Q'/\partial I)_{\text{tot}}$ with field for annealed cobalt. The rate of change of internal magnetic energy with magnetization is $(\partial E_M/\partial I)_S = H-(\partial Q'/\partial I)_S$.

§ 6. CONCLUSION

A series of comprehensive measurements on the magnetic properties of cobalt and nickel has been made. It has been shown that, in high fields, the reversible magneto-thermal effect in nickel is to be associated mainly with changes in the intrinsic magnetization and the rotation of the magnetization vectors against the anisotropy forces. If the marked fall in the numerical value of b'' observed in the earlier total magneto-thermal measurements is real, it must be the result of some irreversible process; there is no indication of any large decrease in the numerical value of b'' in the results of the present measurements. At low fields the results for nickel, which are in general agreement with those in I, indicate that in these fields some additional reversible effect, such as reversible boundary movements, take place. It would appear that irreversible effects are present outside what is normally considered the hysteresis region. The investigations on cobalt have led to the conclusion that, even in the hysteresis region, the only reversible processes involved are rotations and intrinsic magnetization changes. The irreversible cooling effect obtained in the measurements in nickel was not observed in cobalt.

One of the most encouraging features of this work has been the good agreement between the results of the present measurements on the total magneto-thermal effect and the corresponding results of Bates and his co-workers, using a different method of measurement; this is the first occasion on which independent investigations on similar materials have been carried out.

ACKNOWLEDGMENT

The authors wish to express their thanks to Professor E. C. Stoner, under whose direction this work has been carried out, and to the Department of Scientific and Industrial Research for a maintenance grant (R. W. T.).

APPENDIX

An ellipsoid, semi-principal axes a, b, c, orientated in the directions Ox, Oy, Oz, respectively, is considered uniformly magnetized parallel to the x axis with magnetic moment M. The potential Ω at a point (x, y, z) outside the ellipsoid is given by

$$\Omega = \frac{3Mx}{2} \int_{\xi}^{\infty} \frac{dS}{(a^2+S)R_s} \qquad \ldots\ldots(\mathrm{A}\ 1)$$

where $R_s = \{(a^2+S)(b^2+S)(c^2+S)\}^{1/2}$ and ξ is the parameter of the family of ellipsoids confocal with the given ellipsoid, for which $\xi = 0$, (see, for example, Stratton 1941.) With $b = c \ll a$, the component of the field at (x, y, z), parallel to Oy is given by

$$H_y = -\frac{3M}{a^3}\left\{\frac{\partial}{\partial y}\left[x \ln\left(\frac{(a^2+\xi)^{1/2}+a}{\xi^{1/2}}\right) - \frac{a}{(a^2+\xi)^{1/2}}\right]\right\} \qquad \ldots\ldots(\mathrm{A}\ 2)$$

$$= -\frac{3M}{a^3}\left\{\frac{\partial}{\partial y}\left[x \ln\left(\frac{u+1}{(u^2-1)^{1/2}}\right) - \frac{1}{u}\right]\right\} \qquad \ldots\ldots(\mathrm{A}\ 3)$$

with

$$\xi/a^2 = \tfrac{1}{2}\{n^2+m^2-1+[(n^2+m^2-1)^2+4m^2]\} \qquad \ldots\ldots(\mathrm{A}\ 4)$$

where $n = x/a$, $m = y/a$ and $u^2 - 1 = \xi/a^2$, and the sign of the square root is positive. Equation (A 3) then gives

$$-\frac{a^3}{3M} H_y = \frac{mn}{u(u^2-1)(2u^2-m^2-n^2-1)}. \qquad \ldots\ldots(\mathrm{A}\ 5)$$

REFERENCES

BATES, L. F., 1951, *J. Phys. Radium*, **12**, 459.

BATES, L. F., and DAVIS, J. H., 1950, *Proc. Phys. Soc.* A, **63**, 1265.

BATES, L. F., and SHERRY, N. P. R., 1955, *Proc. Phys. Soc.* B, **68**, 642.

CARSLAW, H. S., and JAEGER, J. C., 1947, *Conduction of Heat in Solids* (Oxford : Clarendon Press), p. 83.

HARDY, T. C., and QUIMBY, S. L., 1938, *Phys. Rev.*, **54**, 217.

STONER, E. C., 1945, *Phil. Mag.*, **36**, 803.

STONER, E. C., and RHODES, P., 1949, *Phil. Mag.*, **40**, 481.

STRATTON, J. A., 1941, *Electromagnetic Theory* (New York : McGraw-Hill), p. 257.

TEBBLE, R. S., and CORNER, W. D., 1950, *Proc. Phys. Soc.* B, **63**, 1005.

TEBBLE, R. S., CORNER, W. D., and WOOD, J. E., 1951, *Proc. Phys. Soc.* B, **64**, 753.

TEBBLE, R. S., WOOD, J. E., and FLORENTIN, J. J., 1952, *Proc. Phys. Soc.* B, **65**, 858.

Concentration Changes in an Aerosol

By E. R. WOODING†

Imperial College of Science and Technology, London

MS. received 1st *May* 1956, *and in revised form* 28*th September* 1956

Abstract. The rate of coagulation of an aerosol was calculated by Smoluchowski on the assumption that every collision resulted in the coalescence of two particles. The coagulation rate when there is a potential barrier round the aerosol particles is now calculated and two examples are considered.

§ 1. Introduction

SMOLUCHOWSKI (1917) proposed a relation between the coagulation rate of an aerosol, the concentration n and the radius a of the particles,

$$dn/dt = -8\pi D a n^2. \qquad \ldots\ldots(1)$$

The diffusivity D of particles in Brownian motion was shown by Einstein (1905) to be

$$D = kT/6\pi\eta a,$$

where k is Boltzmann's constant, T the absolute temperature and η the viscosity of the gas.

It was assumed that two colliding particles would always coalesce to form a single one of double mass. If there is a potential barrier around each particle, the proportion of particles penetrating is related to their translational energy. Since the distribution of energy among particles in Brownian motion is known (Chandrasekhar 1943) it is possible to determine the proportion of particles which coalesce and hence to calculate a new coagulation rate.

There is some experimental evidence of the existence of such an effect in aerosols. Swinbank (1947) observed particles of a few microns diameter which were projected at one another, and he reported that they failed to coalesce on collision. Dady (1947) carried out similar experiments. However, work of this type necessitates location of the particle trajectories to within one particle radius if the results are to be significant.

By means of an ultramicroscope Fuchs (1935) succeeded in detecting the rebound of particles after collision with a wall, and observed that 10 to 25% of the particles rebounded.

Tikhomirov *et al.* (1942) claimed that intermolecular forces could affect the rate of coagulation by increasing the probability of coalescence in collisions. They obtained experimental values of the coagulation rates, $8\pi Da$, in various aerosols, which were 1·2 or 1·3 times as large as the values predicted by Smoluchowski.

Tunitsky (1939) derived an expression for the rate of coagulation of electrically charged particles.

It is now proposed to examine a general expression for the rate of coagulation when there is a potential barrier around each particle.

---

† Now at Department of Physical Chemistry, University of Sheffield.

§ 2. A General Expression for the Coagulation of an Aerosol

The rate of change of concentration in an aerosol in which the particles diffuse in a field of force $\mathbf{K}$ per unit mass was derived by Smoluchowski (1915). The equation is discussed in a review of stochastic processes by Chandrasekhar (1943) and is

$$\frac{dn}{dt}=\operatorname{div}_r\left[D\operatorname{grad}_r n-\frac{\mathbf{K}n}{\beta}\right] \qquad \ldots\ldots(2)$$

where $\mathbf{r}$ is the radius vector, $\beta=6\pi\eta a/mf$ is the viscous drag on a particle per unit mass, m is the mass of a particle and f the slip factor.

This equation is now used to derive an expression for the diffusion of particles towards a spherical surface surrounded by a potential barrier. The force $\mathbf{K}$ on the particles is the gradient of the potential V, i.e.,

$$\mathbf{K}=-\operatorname{grad}_r V. \qquad \ldots\ldots(3)$$

Such a force may be caused by free electric charges on the particles or by intermolecular forces.

The diffusion current $\mathbf{j}$ of particles is related to the concentration changes by

$$dn/dt=-\operatorname{div}_r \mathbf{j}. \qquad \ldots\ldots(4)$$

Then equation (2) may be written

$$\mathbf{j}=-[D\operatorname{grad}_r n-\mathbf{K}n/\beta]. \qquad \ldots\ldots(5)$$

Substituting the potential V from equation (3) and rearranging:

$$\mathbf{j}=-D\exp(-V/D\beta)\operatorname{grad}_r[n\exp(V/D\beta)]. \qquad \ldots\ldots(6)$$

The flux F of particles across a surface s normal to the current is related to the current by:

$$\mathbf{j}=\frac{\partial\mathbf{F}}{\partial s}\frac{1}{r^2}=\frac{\partial\mathbf{F}}{\partial\omega}$$

where ω is a solid angle. Hence

$$r^{-2}\partial\mathbf{F}/\partial\omega\exp(\gamma V)=-D\operatorname{grad}_r[n\exp(\gamma V)] \qquad \ldots\ldots(7)$$

(cf. Kramers 1940) where $\gamma=(D\beta)^{-1}$. Since the system is spherically symmetrical, the flux into unit solid angle is independent of r and equation (7) may be integrated to obtain

$$\frac{\partial F}{\partial\omega}=\frac{\int_R^\infty D\operatorname{grad}_r[n\exp(\gamma V)]dr}{\int_R^\infty r^{-2}\exp(\gamma V)dr}$$

where R is the radius of the sphere of influence of a particle. The flux into the whole sphere is

$$F=\frac{\left[4\pi Dn\exp(\gamma V)\right]_R^\infty}{\int_R^\infty r^{-2}\exp(\gamma V)dr}.$$

The boundary condition at $r=R=2a$ is $n=0$ when every particle making contact is absorbed (or coalesces). At a distance from the sphere large compared with R the concentration is n_0 and the potential V is generally zero. Thus

$$F=\frac{4\pi Dn_0}{\int_R^\infty r^{-2}\exp(\gamma V)dr}.$$

When the particle is not stationary, but is itself diffusing, the relative diffusivity D_{12} of the two colliding particles must replace D where

$$D_{12}=\frac{n_1D_1}{n_0}+\frac{n_2D_2}{n_0}=D$$

since the particles are identical. The potential V_{12} is the potential of one particle with respect to the other. The rate of coagulation dn/dt is the flux of particles into n_0 such sinks:

$$\frac{dn}{dt}=\frac{-4\pi Dn_0^2}{\int_{2a}^{\infty} r^{-2}\exp(V_{12}/kT)dr} \qquad \ldots\ldots(8)$$

where γ has been replaced by m/kT.

§ 3. EXAMPLES

Examples will now be considered of coagulating aerosols in which particles must penetrate a potential barrier before they can coalesce.

3.1. *The Coalescence of Particles having an Electric Charge*

For two particles having charges e_1 and e_2, and separated by a distance r, the potential V is given by $V_{12}=e_1e_2/r$ (this is accurate for point charges only). Then equation (8) becomes

$$\frac{dn}{dt}=\frac{-4\pi Dn^2}{\int_{2a}^{\infty} r^{-2}\exp(e_1e_2/rkT)dr}.$$

Now putting $r=1/u$ and $\epsilon=e_1e_2/kT$

$$\frac{dn}{dt}=\frac{-4\pi Dn^2}{\int_0^{a/2}\exp\epsilon u\,du}$$

$$=\frac{-4\pi Dn^2\epsilon}{\exp(\epsilon/2a)-1}$$

(cf. Tunitsky 1939). If $\epsilon/2a$ is small

$$dn/dt\simeq\frac{-8\pi aDn^2}{(1+\epsilon/4a)}. \qquad \ldots\ldots(9)$$

3.1.1. When particles are uncharged, ϵ is zero and $dn/dt=-8\pi Dan^2$ which is the result obtained by Smoluchowski (1917).

3.1.2. When the colliding particles have charges of equal magnitude and opposite sign,

$$e_1=-e_2=e$$

$$\epsilon=-e^2/kT=-\epsilon_1$$

and

$$\frac{1}{1-\epsilon_1/4a}>1.$$

3.1.3. When the particles have charges of equal magnitude and the same sign,

$$e_1 = e_2 = e$$

and

$$\frac{1}{1+\epsilon_2/4a} < 1$$

and the aerosol is more stable than predicted by Smoluchowski's relation.

3.1.4. It can be seen that the expansion used to obtain equation (9) is valid, provided e_1e_2/a is less than about 10, for $\epsilon/4a = 1{\cdot}42(e_1e_2/a)10^{-2}$ when e_1 and e_2 are the number of electron (e.s.u.) charges and a is in microns. When the particle size is of the order of 1 micron, charges of 10 electron units will have an appreciable effect on the rate of coagulation.

Consider an aerosol in which b particles per cubic centimetre have a positive charge, c a negative charge and d $(=n-b-c)$ are uncharged. Each of the positive particles may collide with $b-1$ similarly charged particles or with c negative particles. Hence the possibilities of collisions among the positive particles are $b(b-1)$ or b^2 since $b \gg 1$. The possible number of collisions between positive and negative particles is $2bc$. The number of possibilities of the various types of collisions occurring are given in the table.

The rate of coagulation is then:

$$\frac{dn}{dt} = -8\pi aD\left[\frac{g}{1-\epsilon_1/4a} + \frac{h}{1+\epsilon_2/4a} + i\right] \qquad \text{......(10)}$$

where g and h are the numbers of collisions which take place between particles of unlike and like charge respectively, and i is the number of collisions of uncharged particles with charged particles and with each other.

A collision between a charged and an uncharged particle has been treated as equivalent to one between uncharged particles. Consequently, equation (10) underestimates the rate of coalescence since a charge distribution is induced on an uncharged dielectric sphere in an electric field.

Collisions Possible among the Various Types of Particle

Type of collision	Participating particles	Collisions cm^{-3}
Between unlike charges	b+ve with c−ve	$g=2bc$
Between like charges	Among b+ve and among c−ve	$h=b^2+c^2$
Between uncharged and between charged and uncharged particles	Among d uncharged, b+ve with d uncharged, and c−ve with d uncharged.	$i=d^2+2bd+2cd$
Total	$n=b+c+d$	$n^2=g+h+i$

3.2. *The Effect of Van der Waals Forces*

The Van der Waals forces produce a mutual attraction between two atoms which has been shown by London (1930) to be equivalent to a potential V where

$$V = -\tfrac{3}{4}(P^2h\nu)/r^6.$$

Here P is the polarizability of the atoms, h Planck's constant and ν a frequency of the order of 10^{15} to 10^{16} c/s, which is attributed to the electrons of neutral atoms generating oscillating dipoles. Consequently $h\nu$ represents a characteristic energy.

Hamaker (1937) discussed the interaction of two spheres and from his result it can be shown that the energy of attraction is

$$V=\frac{-B}{12}\left[\frac{4a^2}{r^2-4a^2}+\frac{4a^2}{r^2}+2\ln\left(\frac{r^2-4a^2}{r^2}\right)\right] \qquad \ldots\ldots(11)$$

where

$$B=\tfrac{3}{4}\pi^2P^2h\nu q^2 \qquad \ldots\ldots(12)$$

and q = number of atoms per cubic centimetre. When the distance $r-2a$ between the surfaces of the two spheres is very small, then equation (11) becomes

$$V=\frac{-Ba}{12(r-2a)} \quad \text{as} \quad \frac{r-2a}{a}\to 0. \qquad \ldots\ldots(13)$$

Substituting this potential into equation (8), the rate of coagulation of neutral particles becomes

$$\frac{dn}{dt}=-\frac{4\pi Dn^2}{\int_{2a+\delta}^{\infty} r^{-2}\exp\left[-C/(r-2a)\right]dr}$$

where $C=Bam/12kT$, δ is a factor representing surface roughness and it is assumed that coalescence occurs when the surfaces of the two spheres are separated by a very small distance of dimension comparable with molecular diameters, i.e. of order 10^{-8} cm. Now putting $x=r/(r-2a)$

$$\begin{aligned}\int_{2a+\delta}^{\infty}\frac{1}{r^2}\exp\left(\frac{-C}{r-2a}\right)dr &= \frac{1}{2a}\exp\left(\frac{C}{2a}\right)\int_{1}^{(2a+\delta)/\delta}\frac{1}{x^2}\exp\left(\frac{-Cx}{2a}\right)dx\\ &\simeq\frac{1}{2a}\left[1-\frac{\delta}{2a+\delta}\exp\left(\frac{-C}{\delta}\right)\right.\\ &\qquad\left.-\frac{C}{2a}\exp\left(\frac{C}{2a}\right)\left(\ln\frac{2a+\delta}{\delta}-\frac{C}{\delta}\right)\right]\\ &=\frac{1}{2a}A\ (\text{say}).\end{aligned}$$

Hence $$dn/dt=-(8\pi Dan^2)/A.$$

From figures given by London (1937) a value 5×10^{-12} for B can be obtained for the interaction between water molecules (equation (12)). Considering water droplets of 1 micron radius, $C=0{\cdot}4\times10^{-15}$ cm so $A=1$ to within 10^{-2}% and the effect of the Van der Waals' forces on coagulation is negligible.

§ 4. CONCLUSION

The theoretical relation (8) which has been derived may be used in the determination of coagulation rates when the number of collisions resulting in coalescence is diminished by various potential barriers. The substitution of numerical values into the equations for the rate of coagulation indicates that an electrically charged aerosol coagulates at a rate which is appreciably different from that of an uncharged one. The effect of Van der Waals forces on coagulation of an aerosol of water droplets is negligible.

Only first order effects have been considered, and changes in particle size distribution during coagulation have yet to be determined. Consequently the coagulation rates derived from equation (8) are subject to an error which increases with the time after formation of the aerosol. There are yet other phenomena which may affect the coagulation rates. Particles of volatile material set up a potential barrier by virtue of the momentum of the evaporating molecules. Hopper and Grant (1948) have shown that the viscous resistance to a drop falling under gravity is appreciably increased at a distance of less than about 10 radii from a solid horizontal surface and it is expected that the effective diffusivities of two particles in close proximity are diminished. Even when contact is made adsorbed films may inhibit the wetting of one particle by another and prevent coalescence.

Acknowledgment

The author is greatly indebted to the British Shipbuilding Research Association who sponsored the work. He is very grateful to Professor D. M. Newitt for his supervision, to Dr. J. F. Richardson for much encouragement and advice and to Dr. C. N. Davies for many helpful comments.

References

Chandrasekhar, S., 1943, *Rev. Mod. Phys.*, **15**, 1.
Dady, G., 1947, *C. R. Acad. Sci., Paris*, **225**, 1349.
Einstein, A., 1905, *Ann. Phys., Lpz.*, **17**, 549.
Fuchs, N., 1935, *Acta Phys. Chim., U.R.S.S.*, **3**, 819.
Hamaker, H. C., 1937, *Physica*, **4**, 1058.
Hopper, V. D., and Grant, A. M., 1948, *Aust. J. Sci. Res.*, **A1**, 28.
Kramers, H. A., 1940, *Physica*, **7**, 284.
London, F., 1930, *Z. Phys.* **63**, 245; 1937, *Trans. Faraday Soc.*, **33**, 8.
Smoluchowski, M., 1915, *Ann. Phys., Lpz.*, **48**, 1103; 1917, *Z. Phys. Chem.*, **92**, 129.
Swinbank, 1947, *Nature, Lond.*, **159**, 849.
Tikhomirov, M. V., Tunitsky, N. N., and Petrjanov, J. B., 1942, *Acta. Phys.-Chim., U.R.S.S.*, **17**, 185.
Tunitsky, N. N., 1939, *J. Phys. Chem., Moscou*, **13**, 1141.

Ultrasonic Relaxation in Methylene Chloride

By J. H. ANDREAE

Imperial Chemical Industries, Limited, Akers Research Laboratories, Welwyn, Herts

Communicated by J. J. Benbow; MS. received 3rd September 1956

Abstract. Ultrasonic absorption measurements have been made in liquid methylene chloride between 7·5 and 210 Mc/s. These confirm recent work in the vapour by Sette, Busala and Hubbard (1955), from which it was concluded that the relaxation of the vibrational specific heat takes place in two parts. In the frequency range studied here there is a single relaxation of specific heat equal in magnitude to the vibrational specific heat less the contribution of the lowest frequency mode of vibration. The relaxation frequency is 170 Mc/s at 25°c.

§ 1. Introduction

Measurements of ultrasonic velocity in gases have been used to study the transfer of translational energy to molecular vibrations. Lambert *et al.* (Lambert and Rowlinson 1950, Fogg, Hanks and Lambert 1953, Fogg and Lambert 1955) have shown for a large number of organic vapours that the ultrasonic dispersion resulting from the relaxation of this transfer of energy may be described, in each case, by a single relaxation time and that, in each case, the whole of the vibrational specific heat is involved. Recent measurements of the ultrasonic absorption in liquid carbon disulphide (Andreae, Heasell and Lamb 1956) have shown that here also there is a single relaxation of the total vibrational specific heat. In methylene chloride, however, Sette, Busala and Hubbard (1955, to be referred to as SBH) have reported velocity measurements which indicate that there are two relaxation regions in the vapour; they attribute the higher frequency region to a relaxation of the vibrational mode with the lowest spectroscopic frequency, while in the other region the remaining eight modes of the molecule relax together.

In the present work ultrasonic absorption measurements in liquid methylene chloride are reported which confirm the conclusions reached by SBH for the vapour.

§ 2. Experimental

A relaxation in liquid methylene chloride was found by the author in 1953 (Andreae 1955), but the ultrasonic absorption measurements taken at that time have been rejected because of water impurity in the liquid samples. The liquid used for the present work was dried over phosphorus pentoxide and distilled, the fraction boiling at 39·3°c (738 mm pressure) being collected. The absorption measurements taken with these samples have yielded values as much as 20% higher than the 1953 values.

All measurements were taken by the pulse technique (Pellam and Galt 1946, Pinkerton 1949, Andreae, Heasell and Lamb 1956). It is sufficient to say that two separate X-cut quartz crystals were employed for generating and detecting the ultrasonic waves. At the lowest frequency (7·5 Mc/s) the crystals were in direct contact with the liquid, but at the higher frequencies the ultrasonic path was very short and fused quartz delay bars were interposed between the crystals and the liquid. In all cases an E_{01} mode piston attenuator was the standard of attenuation.

§ 3. RESULTS AND CALCULATIONS

The experimental values for the ultrasonic absorption in methylene chloride are given in the first three columns of table 1. At each frequency an average of two or more measurements has been taken. The scatter is greater at the higher frequencies, but an accuracy to ±2% is claimed for the apparatus. This does not cover the possibility of higher values being obtained with even purer samples of the liquid.

Table 1. Ultrasonic Absorption in Methylene Chloride

Experimental			Calculated		
Frequency f (Mc/s)	Temperature (°c)	$10^{17}\,/\alpha f^2$ (sec^2 cm^{-1})	Velocity (m sec^{-1})	$10^{17}\,\alpha/f^2$ (sec^2 cm^{-1})	Difference (%)
7·54	25·2	1220	1050	1222	+0·2
50·3	25·1	1120	1053	1128	+0·7
70·0	25·0	1060	1055	1053	−0·7
110	25·2	875	1064	875	0
169	25·0	640	1067	645	+0·8
209	24·9	530	1071	526	−0·8

Calculated values are given for $A=1200\times10^{-12}$, $B=85\times10^{-12}$ second, and $f_a=170$ Mc/s in equation (1).

Preliminary calculations show that the experimental values correspond to a single relaxation process and it will be assumed that this process is the relaxation of vibrational specific heat. However, the total vibrational specific heat of methylene chloride at 25°c is too large to account for the ultrasonic measurements.

If only part of the vibrational specific heat is contributing to the relaxation this part may be found by fitting the experimental values to a curve of the form (Andreae and Lamb 1956)

$$\frac{\alpha c}{f^2}=\frac{A}{1+(f/f_a)^2}+B, \qquad \ldots\ldots(1)$$

where α is the ultrasonic absorption coefficient, c is the phase velocity and f is the frequency. A, B and f_a are parameters to be adjusted in the curve-fitting. To begin with, the phase velocity is assumed to be independent of frequency and then, when an approximate value of the relaxing specific heat C_{1a} has been calculated from the formula

$$C_{1a}=\frac{Af_a}{\pi}\cdot\frac{C_{p0}}{\gamma-1}\left(\frac{C_{p1}C_{v1}}{C_{p0}C_{v0}}\right)^{1/2}, \qquad \ldots\ldots(2)$$

allowance can be made for the small velocity dispersion by means of the formula

$$(c/c_0)^2=1+\left(1-\frac{1}{\gamma}\right)\frac{C_{1a}}{C_{v1}}\,\frac{1}{1+(f_a/f)^2}. \qquad \ldots\ldots(3)$$

These two formulae are the simpler 'thermal relaxation' versions of the more general relations given for 'structural relaxations' by Andreae and Lamb (1956). The subscripts 0 and 1 on the symbols C_p and C_v for the specific heats at constant pressure and volume indicate that the symbols refer to values at frequencies respectively much lower and much higher than the relaxation frequency f_a. γ is the ratio C_{p0}/C_{v0}. On the reasonable assumption that there are no other processes in the liquid relaxing at frequencies below 7·5 Mc/s, C_{p0} and C_{v0} are identified with the static specific heats of the liquid, and C_{p1} and C_{v1} are obtained by subtracting the relaxing specific heat from C_{p0} and C_{v0}, viz. $C_{p1}=C_{p0}-C_{1a}$, $C_{v1}=C_{v0}-C_{1a}$. But C_{1a} is not yet known so a process of trial and error is employed which gives values which converge rapidly because the ratios under the square root in equation (2) are insensitive to changes in the value of C_{1a}.

This procedure results in the values given in table 2, where the other data for methylene chloride are also listed. The agreement with experiment is shown by the figures in the fourth, fifth and sixth columns of table 1 which are obtained from formulae (1), (2) and (3) with the values shown in table 2. A variation of 3% in the values of A and f_a does not spoil this agreement significantly.

Table 2. Data for Methylene Chloride at 25°c

Thermal expansion coefficient†	θ	$\text{deg}^{-1}\times 10^3$	1·367
Specific heat at constant pressure†	C_{p0}	cal mole^{-1} deg^{-1}	24·18
Ultrasonic velocity‡	c_0	m sec^{-1}	1050
Ratio of specific heats	γ§		1·516
	A	$\text{sec}\times 10^{-12}$	1200
	B	$\text{sec}\times 10^{-12}$	85
	f_a	Mc/s	170
	C_{1a}	cal mole^{-1} deg^{-1}	2·63

† Timmermans 1950; ‡ Lagemann, McMillan and Woolf 1949.

§ $\gamma = C_{p0}/C_{v0} = (1+\theta^2 T\,{}_0^2/C_{p0}$

§ 4. DISCUSSION

The vibrational specific heat of methylene chloride is composed of the contributions from the nine modes of vibration. The vibration frequencies ν are listed in table 3, together with the components of the specific heat calculated from the Planck–Einstein formula

$$C_1 = \boldsymbol{R}x^2\frac{e^{-x}}{(1-e^{-x})^2} \qquad \ldots\ldots(4)$$

where $x=\boldsymbol{h}\nu/\boldsymbol{k}T$.

The total vibrational specific heat is seen to be 4·29 cal mole^{-1} deg^{-1}, of which only 2·63 cal mole^{-1} deg^{-1} (table 2) appear to take part in the ultrasonic relaxation. The velocity measurements of SBH indicate that there are two relaxation regions in the vapour; they attribute the higher frequency region to the lowest mode of vibration by itself, while in the lower region the remaining eight vibrational modes of the methylene chloride molecule relax together. If the relaxation in the liquid is assumed to correspond to the lower frequency relaxation in the vapour, then the conclusions of SBH are confirmed by the measurements in the liquid: the relaxing specific heat (2·63 cal mole^{-1} deg^{-1}) is practically equal to the total vibrational specific heat minus the specific heat contributed by the lowest mode ($4{\cdot}29-1{\cdot}70=2{\cdot}59$ cal mole^{-1} deg^{-1}).

There are, however, other ways in which the vibrational specific heat may be divided into two parts. For instance, it has sometimes been assumed that only the lower energy levels of all the modes contribute to the relaxing specific heat, but in this case it is difficult to understand why any particular level should separate the contributing part from the non-contributing part. It is easier to accept a partition of the modes into two groups, especially if it can be shown that certain modes of vibration are more likely to interact with each other than with the remainder. The coincidence of harmonic frequencies of two modes of vibration may lead to a Fermi resonance and a coupling of energy between these modes, so long as the vibrations are of the same species. It is worth mentioning that the third (849 cm^{-1}), fourth (1132 cm^{-1}) and fifth (1415 cm^{-1}) harmonic frequencies of ν_1 (283 cm^{-1}) in table 3 correspond roughly to the frequencies ν_4, ν_5 and ν_7 respectively, and the sum of the contributions to the specific heat from these four modes amounts to 2·54 cal mole^{-1} deg^{-1}, which approaches the required value of 2·63 cal mole^{-1} deg^{-1}. However, Herzberg (1945) does not assign these four vibrations the same species so a Fermi resonance should not be possible. Four vibrations which he does assign the same species are ν_1, ν_2, ν_7 and ν_8, and the sum of the specific heat contributions from these modes amounts to 2·62 cal mole^{-1} deg^{-1}, which is almost identical with the value obtained from the ultrasonic measurements. It is not difficult to see that harmonic resonance could be invoked here also, though in less simple a manner.

A final decision on the accuracy of SBH's explanation of the partition of the vibrational specific heat should be possible when low temperature measurements are available. The temperature variation of the relaxing specific heat will be markedly affected by the presence or otherwise of the contribution of the lowest mode. For instance, at −60°c the contribution of this mode is only 13% lower than at +25°c, while all the contributions of the other modes have decreased by more than 50%. If the partition is due to Fermi resonance ($\nu_1+\nu_4+\nu_5+\nu_7$ or $\nu_1+\nu_2+\nu_7+\nu_8$), as discussed above, the contribution of the lowest mode will be present and the relaxing specific heat will not change much with temperature. On the other hand, if the lowest mode is absent, as suggested by SBH, then the relaxing specific heat will diminish rapidly as the temperature falls.

Table 3. Vibrational Specific Heat of Methylene Chloride at 25°c

ν	cm^{-1}	$h\nu/kT$	C_i (cal mole^{-1} deg^{-1})	ν'
1	283	1·37	1·704	4
2	704	3·40	0·821	3
3	737	3·56	0·759	9
4	899	4·34	0·501	7
5	1155	5·57	0·237	5
6	1266	6·11	0·168	8
7	1429	6·90	0·097	2
8	2984	14·4	negligible	1
9	3048	14·7	negligible	6
			4·29	

The spectroscopic frequencies are taken from Herzberg (1945) and the column under ν' gives his nomenclature. The numbers referred to in the text are in the column under ν.

Fogg and Lambert (1955) have shown for a large number of organic vapours that ultrasonic dispersion resulting from the relaxation of the vibrational specific heat may be described, in each case, by a single relaxation time and that, in each case, the whole of the vibrational specific heat is involved. Following the theoretical treatment of Schäfer (1940), they have found that the relaxation times are related to the spectroscopic frequencies of the lowest modes of vibration. This is explained in terms of a series excitation by which energy is first transferred from translational motion to the lowest frequency mode and then is rapidly distributed amongst the higher modes. For each vapour Fogg and Lambert plot $\log Z_{10}$ against the spectroscopic frequency of the lowest vibrational mode. Z_{10} is the number of intermolecular collisions necessary for a molecule to lose one quantum of vibrational energy and it is proportional to the relaxation time. All the points for the vapours measured lie approximately on one of two straight lines emanating from the origin. Fogg and Lambert show that there is a tendency for molecules whose lowest vibrations are infra-red inactive, or only weakly active, to lie along the 'longer relaxation time line', while those whose lowest vibrations are infra-red active lie along the other 'shorter relaxation time line'.

The rate at which energy is transferred from translational to vibrational motion is, therefore, partly due to the infra-red activity of the lowest frequency vibration. Exceptions to the rule show that this is not always the case. It has already been seen that in methylene chloride the lowest frequency vibration probably acts separately and so the second lowest vibration (ν_2) becomes, effectively, the lowest one in the transfer of energy to the higher frequency vibrations. Only when this is assumed does methylene chloride lie with methylene fluoride on Fogg and Lambert's shorter relaxation time line. Other exceptions are carbon dioxide, carbon disulphide, carbonyl sulphide, nitrous oxide, benzene and cyclopropane. Fogg and Lambert suggest that steric effects play a predominant role here because all the molecules are linear or planar.

Among liquids, carbon disulphide (Andreae, Heasell and Lamb 1956) has the lowest known vibrational relaxation frequency, 78 Mc/s, methylene chloride is next with 170 Mc/s and from their high value of ultrasonic absorption it is likely that benzene and carbon tetrachloride will be at frequencies not much higher. Of these four liquids only methylene chloride has a molecule with a dipole moment, and so it may be deduced that the exchange of translational and vibrational energy is slower with non-polar liquids. The unique position here of methylene chloride, which has an appreciable dipole moment, is accounted for by the double relaxation discussed earlier. Since the lowest mode of vibration probably relaxes at much higher frequencies, the relaxation in the region of 170 Mc/s corresponds to a molecule with a lowest vibration frequency equal to 704 cm^{-1} (ν_2): this is the second lowest vibration frequency of methylene chloride. According to Fogg and Lambert's results the relaxation time increases with the spectroscopic frequency of the lowest vibration and so methylene chloride has a relaxation frequency much less than other liquids with lower spectroscopic frequencies for their lowest vibrations. In fact, apart from water which is highly associated, most substances which exist as liquids at room temperature have vibration frequencies below 600 cm^{-1}. If these arguments are valid, ultrasonic measurements on some liquefied gases may reveal vibrational relaxations at frequencies below 200 Mc/s.

Table 3 shows that there is a large gap in frequency between ν_1 (283 cm^{-1}) and ν_2 (704 cm^{-1}). A brief survey of available data did not reveal any other molecule with such a large difference in frequency between the lowest and the next lowest mode. It is possible that this is the reason for the separate relaxation of the lowest vibrational mode of methylene chloride. ν_2 is 2·5 times as large as ν_1 so that, if some kind of harmonic coupling is responsible for the distribution of energy amongst vibrational modes in a molecule, weak coupling could be expected between ν_1 and ν_2.

Intermolecular forces probably play a more important part in liquids than in gases, but the ultrasonic measurements in liquid methylene chloride indicate that even in liquids the spectroscopic frequency of the lowest mode of vibration is a determining factor in the rate of transfer of translational energy to molecular vibrations.

REFERENCES

ANDREAE, J. H., 1955, *Ph.D. Thesis, University of London.*

ANDREAE, J. H., HEASELL, E. H., and LAMB, J., 1956, *Proc. Phys. Soc.* B, **69**, 625.

ANDREAE, J. H., and LAMB, J., 1956, *Proc. Phys. Soc.* B, **69**, 814.

FOGG, P. G. T., HANKS, P. A., and LAMBERT, J. D., 1953, *Proc. Roy. Soc.* A, **219**, 490.

FOGG, P. G. T., and LAMBERT, J. D., 1955, *Proc. Roy. Soc.* A, **232**, 537.

HERZBERG, G., 1945, *Infra-red and Raman Spectra of Polyatomic Molecules* (New York: Van Nostrand).

LAGEMANN, R. T., MCMILLAN, D. R., and WOOLF, W. E., 1949, *J. Chem. Phys.*, **17**, 369.

LAMBERT, J. D., and ROWLINSON, J. S., 1950, *Proc. Roy. Soc* A, **204**, 424.

PELLAM, J. R., and GALT, J. K., 1946, *J. Chem. Phys.*, **14**, 608.

PINKERTON, J. M. M., 1949, *Proc. Phys. Soc.*, **62**, 286.

SCHÄFER, K., 1940, *Z. Phys. Chem.* B, **46**, 212.

SETTE, D., BUSALA, A., and HUBBARD, J. C., 1955, *J. Chem. Phys.*, **23**, 787.

TIMMERMANS, J., 1950, *Physico-Chemical Constants of Pure Organic Compounds*, (New York: Elsevier).

Surface Structures and Ferromagnetic Domain Sizes

By D. H. MARTIN†

Department of Physics, University of Nottingham

Communicated by L. F. Bates; *MS. received* 31*st August* 1956

Abstract. Studies of the domain structures in the surface layers of crystals of iron and silicon–iron having (111) and ($h\,0l$) type faces, have permitted calculations to be made of the optimum domain sizes in iron crystals as dependent on the crystallographic orientations of their faces and on their sizes.

---

§ 1. Introduction

As pointed out by Landau and Lifshitz (1935) subsidiary domains form in the surface layers of crystals of ferromagnetic metals since the total magnetic energy of the crystal is thus reduced. Moreover, the sizes of the *main domains* inside the crystal are controlled by the energy density associated with the surface structures (for a review of the subject see Kittel 1949). Many magnetic properties must be dependent upon the volume of the main domains, e.g. susceptibility, and yet studies have been made hitherto only of the surface structures which form at surfaces of specially simple crystallographic orientations. The work recorded here consists firstly of a powder pattern examination of the surface structures which form in large crystals of iron at (111) and ($0\,kl$) type surfaces, and secondly of calculations of optimum main domain sizes using models of domain structures suggested by the powder patterns. The dependence of domain size upon a crystal's face-orientations is found to be the more marked the smaller the crystal, but is, perhaps, less pronounced than might have been expected *a priori*.

Landau and Lifshitz (1935) suggested, for a demagnetized crystal, the model of surface structures shown in figure 1 (*a*). The surface domains are magnetized in the plane of the surface at which they form and thereby give rise to zero magnetostatic energy. This will certainly be a structure of minimum free energy so long as the direction in which the surface domains are magnetized is an easy direction of magnetization, i.e. of the type [100] in iron. Surface domains of this type have in

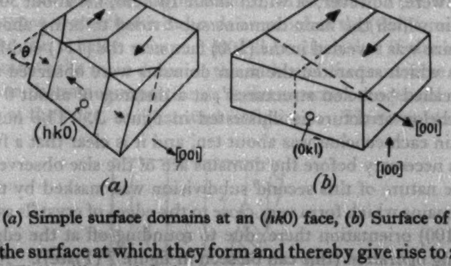


Figure 1. (*a*) Simple surface domains at an ($hk0$) face, (*b*) Surface of ($0kl$) type.

---

† Now at Queen Mary College, University of London.

fact already been observed (Bozorth 1951) in crystals of silicon–iron with (010) faces, and also in similar crystals of pure iron during the present work. If the crystal face contains no easy direction of magnetization, however, the structure of figure 1 (*a*) would be associated with a large magneto-crystalline energy and either a severe reduction in domain size is to be expected or a change in the type of surface domain. The studies reported in § 2 are concerned with the surface domains at (111) faces, which in iron contain no easy direction, and at (0*kl*) faces which contain only one easy direction. The calculations in § 3, based on these studies, lead to conclusions about the domain sizes in crystals bounded by faces of arbitrary orientation, and the influence of the crystal dimensions.

The powder pattern technique for studying domain structures consists of applying a liquid suspension of very fine particles of magnetite to a strain-free electropolished surface. The stray fields there, arising from the underlying domain structure, cause the magnetite particles to form patterns on the surface. These may be observed through a microscope and, provided the orientation of the surface has been suitably chosen, they reveal the underlying domains (cf. Bates and Mee 1952). The crystals used in the work below were prepared with (100) faces as well as the faces at which the structures of interest formed. Since the (100) faces contained the easy directions of magnetization in which the main domains were magnetized they required no surface structures and the *main* domains were directly revealed in them. The surface structures to be studied formed in the *end-faces* of these crystals and the subdivision of the main domains to form them was observed in the (100) faces.

§ 2. Experimental Results

2.1. *Surfaces of* (111) *Orientation*

A crystal of pure iron was grown by the strain-anneal process in a bar of Special Pure Swedish Iron which was kindly supplied by the British Iron and Steel Research Association. It was prepared with four (100) type faces and two (111) end-faces. The powder pattern observed on the (111) face is shown in the photograph of figure 2 (Plate). This reveals a complicated structure involving domains of width about 3×10^{-4} cm. The main domains in the crystal, observed in the (100) surface, were, however, of width about 10^{-1} cm, i.e. about 300 times larger. The manner in which the main domains subdivided to bring about this marked difference in size was revealed in the (100) face near the (111) end-faces.

180° walls which separated the main domains were observed to branch into what will be called 'echelon structures', at a distance of about 0·1 cm from the edge. An echelon structure is illustrated in figure 3. The number of small domains within each echelon was about ten, and it is clear that a further stage of subdivision is necessary before the domains are of the size observed on the (111) surface. The nature of this second subdivision was masked by the small spiky superficial domains which form very close to the edges of crystals as a result of the loss of true (100) orientation there, due to rounding-off at the edges during the electropolishing operation (this can be seen in figure 7 (Plate)). A knowledge of the form of such superficial 'tree' domains (see Williams, Bozorth and Shockley 1949), however, permitted the recognition, at several points, of the two types of structure illustrated in figure 4, and repeated subdivisions of these kinds presumably effect the reduction in domain size by the required factor of about 30.

2.2. *Surfaces of the Types* (*hk*0), (0*kl*)

A 180° wall must lie in a plane containing the directions of magnetization of those domains which it separates. If the plane is near an (001) orientation (with reference to figure 1) then for crystal faces with (*hk*0) orientations, i.e. containing the [001] direction, simple surface domains of the type suggested by Landau and Lifshitz are to be expected as illustrated in figure 1 (*a*). When the angle θ is less than about 5° the surface domains degenerate into the well-known tree-structures (Williams, Bozorth and Shockley 1949).

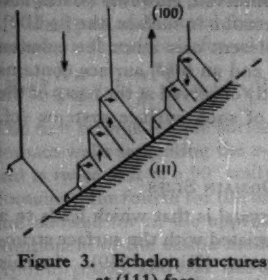


Figure 3. Echelon structures at (111) face.

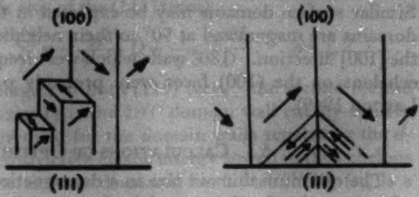


Figure 4. Further subdivision of echelon domains.

If the surface has an orientation (0*kl*) as in figure 1 (*b*), i.e. containing the [100] direction but not the [001], then, for a *main* domain wall near an (001) orientation, the simple surface structures magnetized in the plane of the surface would involve a large magneto-crystalline energy. A less simple structure may, therefore, be expected.

A thin disc was cut from a single crystal of 3% Si–Fe. Its main face was of (100) orientation and its edges were, therefore, of the (0*kl*) family. The powder patterns formed near these edges were observed. Echelon subdivisions were again revealed, and a photograph is shown in figure 5 (Plate). The echelon structures contain two types of domain, that with magnetization transverse to main domains and that with parallel magnetization. The ratio of the widths of the domains of these types in an echelon depends on which of the (0*kl*) family is parallel to the edge where the echelon occurs. It will be clear that if the echelons are uniformly to terminate the main domains the widths of the transverse and parallel domains must be in the same proportion as their respective tangential components of magnetization at the edge, as illustrated in figure 6. This was observed to be

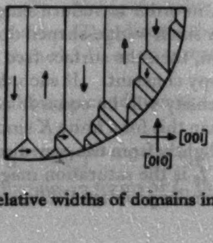


Figure 6. Relative widths of domains in the echelons.

approximately so, as, for example, in the case shown in the photograph of figure 7. In particular the echelon structures degenerated into the simple surface domains with triangular cross sections where the edge's orientation was close to (010). The crystal was too thin to permit powder pattern observations on the (0*kl*) faces themselves. Surface domains in a (101) face of a crystal containing *main* domains magnetized at 90° to their neighbours have been observed, however (Bates and Mee 1952), and were magnetized in the only easy direction in surface, the [± 100]. Similar surface domains may be expected in the present case since the echelon domains are magnetized at 90° to their neighbours and an (0*kl*) surface contains the [100] direction. 180° walls which were frequently observed at the bases of the echelons on the (100) faces were probably parts of such surface systems (cf. Lawton 1949).

§ 3. CALCULATIONS OF OPTIMUM DOMAIN SIZES

The optimum domain size in a demagnetized crystal is that which leads to a minimum value for the sum of the free energies associated with the surface structures and with the walls separating the main domains (Kittel 1949). In this section calculations are made of optimum domain sizes using models of domain structures based on the observations recorded in the previous section, and the results for crystals bounded by surfaces of various orientations are compared over the whole range of crystal sizes.

The first model, for a crystal bounded by (111) faces, is illustrated in figure 8. It is made up of three major regions characterized by domains of widths *D*, *d* and *a* respectively.

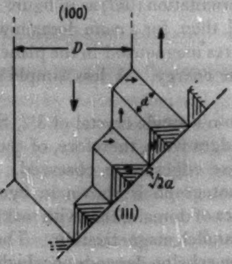


Figure 8. Model of domain structure for calculations of § 3.

The magnetic energy of the (111) surface will depend on what happens where the *a*-domains meet it. If, in order to reduce magnetostatic energy to zero, the *a*-domains are terminated by small wedge-shaped domains magnetized in the (111) surface in the [011] direction, then the surface free energy density will be $\frac{1}{4}K\,.\,\frac{1}{4}a$ where K is the first anisotropy constant. If such domains do not form, the magnetostatic surface energy density will be equal to cI_s^2a, where c would be 1/3 if K were infinite, but is less than this for a finite K since the magnetization very close to the surface may deviate slightly from the easy directions under the action of the local magnetostatic fields. I_s is the saturation magnetization. The value of c is difficult to calculate. In a similar problem Williams, Bozorth and Shockley (1949) showed that the magnetostatic energy of a certain structure was 25 times smaller in iron than it would have been in a material with an infinite K. Such a factor in the present case would make the surface energies of the two possibilities considered above about equal. The patterns observed on (111) faces do not support the one possibility more than the other, but the energy density must, in any case, be about $\frac{1}{4}K \,.\, \frac{1}{4}a$—certainly no more—and this expression will be used.

That part of the total free energy which depends on the sizes of the various domains is therefore (referred to unit area of the (111) faces)

$$F=2\frac{K}{4}\frac{a}{4}+2\gamma\frac{a}{d}\sum_{1}^{d/\sqrt{2}a} n+2\sqrt{2}\frac{\gamma}{2}\frac{d}{D}\sum_{1}^{D/\sqrt{2}d} n+\frac{c_{11}\lambda^2}{4\sqrt{3}}D+\frac{\gamma}{\sqrt{3}}\frac{\sqrt{3}L}{D}. \qquad n=1,2,3\ldots$$

The first term is for the surface energies of both (111) faces; the second for the domain walls separating the a-domains (γ is the 180° domain wall energy density and is twice that of a 90° wall); the third for the domain walls separating the d-domains; the fourth for the magnetostrictive elastic energy stored in the echelons and allowed for in the approximate manner discussed by Kittel (1949) (c_{11} is a component of the elastic modulus and λ is the magnetostriction constant for the [100] direction, and $c_{11}\lambda^2$ has the value $10^3\,\mathrm{erg\,cm^{-3}}$ for iron); and the last term is for the domain walls separating the main domains, where L is the perpendicular distance between the (111) faces.

The optimum domain sizes are given by the three equilibrium conditions:

$$\left(\frac{\partial F}{\partial a}\right)_{d,D}=0; \quad \left(\frac{\partial F}{\partial d}\right)_{a,D}=0; \quad \left(\frac{\partial F}{\partial D}\right)_{a,d}=0.$$

The problem of finding the optimum value of D as a function of L is most easily solved by taking these conditions in the above order.

The first condition involves only the first two terms above and gives the optimum width of the a-domains in terms of d:

$$a=\left(\frac{8d\gamma}{K}\right)^{1/2} \qquad \ldots\ldots(1)$$

A comparison may here be made with the observed domains discussed in § 2, where d was seen to be about 10^{-2} cm. The relation (1) above, with this value for d, gives $a=4{\cdot}1\times10^{-4}$ cm taking $\gamma=1{\cdot}0\,\mathrm{erg\,cm^{-2}}$ and $K=5\times10^{5}\,\mathrm{erg\,cm^{-3}}$, values roughly appropriate for iron. The observed value of a was about 3×10^{-4} cm and, therefore, shows satisfactory agreement with equation (1).

Using equation (1) the second equilibrium condition gives

$$d=\left(\frac{\gamma D^2}{K}\right)^{1/3}. \qquad \ldots\ldots(2)$$

A complete treatment of the magnetostrictive energy of the echelons would no doubt reveal a dependence on d unlike the approximate expression used above, and this fact would modify equation (2). Using the value 0·1 cm for D as observed in the crystal described in § 2, equation (2) indicates that there should be about 40 d-domains per echelon rather than the 10 or so recorded. In the circumstances this is not a serious discrepancy.

Application of the third equilibrium condition and the use of equations (1) and (2) lead to

$$L=0{\cdot}352\,(\gamma^{-1}KD^4)^{1/3}+0{\cdot}144\,c_{11}\lambda^2\gamma^{-1}D^2 \qquad \ldots\ldots(3)$$

giving D, the width of a main domain, in terms of L, the distance between the (111) faces. This relation is plotted in figure 9, curve A, for which the values of the various constants quoted earlier for iron have been used.

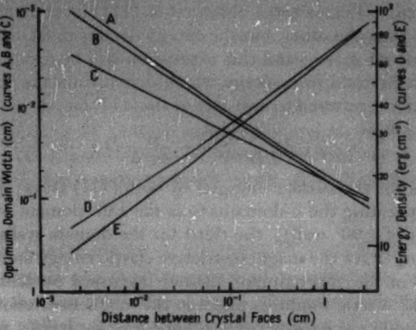


Figure 9. Curves A and B show the dependence of domain width, in an iron crystal bounded by (111) faces, on the distance between the faces; A is for domains separated by 180° walls and B is for domains separated by 90° walls. Curve C shows the variation of domain width in a crystal bounded by (010) faces. Curves D and E show the free energy per unit area of the (111) plane for crystals bounded by (111) surfaces and containing main domains separated by 180° and 90° walls respectively.

The crystal described in § 2.1. was about 1·0 cm in length in the [010] direction, and the relation above therefore predicts a domain width of 0·056 cm, giving satisfactory agreement, in a problem of this nature, with the observed domain width of about 0·1 cm.

The optimum width for domains in a crystal bounded, not by (111) faces as above, but by the (010) faces, can readily be calculated. In such a crystal simple surface domains form of the type shown in figure 1. The energy associated with these domains is mainly magnetostrictive elastic energy, and that part of the free energy dependent on the width Δ of the main domains, in a crystal whose (010) faces are separated by a distance Λ, is (referred to unit area of the (010) surface)

$$F = 0{\cdot}25\, c_{11} \lambda^2 \Delta + \gamma \Lambda \Delta^{-1}.$$

The equilibrium condition $(\partial F/\partial \Delta)_{\Lambda} = 0$ gives

$$\Lambda = 0{\cdot}25 \frac{c_{11}\lambda^2}{\gamma} \Delta^2. \qquad \ldots\ldots(4)$$

This dependence of Δ and Λ is plotted in figure 9, curve C.

The magnetostrictive elastic energy appears in both the (111) and (010) surface structures. The additional (domain wall) energy in the case of the (111) structure leads to a smaller domain size in crystals with (111) faces than in those with (010) faces. Figure 9 shows, however, that the difference in domain width is only about 12% for crystals as large as 1·0 cm in face separation. The difference increases for smaller crystals, being a factor 2 for a face separation of 5×10^{-2} cm and a factor 3 at 5×10^{-3} cm.

When calculations for crystals which are equal in length in the [010] direction, rather than in face separation, are compared, the difference between the domain widths in (111) and (010) crystals is more marked, being a factor 4·5 at 5×10^{-3} cm and about 1·5 at 1·0 cm. Crystals bounded by faces of intermediate orientations, of the ($h0l$) types, have been described in § 2.2. Their surface energies are mainly magnetostrictive and the consequent optimum domain sizes therefore differ little from those in the simple (010) crystal.

No mention has been made so far of *main* domains separated not by 180° walls but by 90° walls. Such walls are not uncommon in most types of crystal. An idealized demagnetized state involving mainly 90° walls is illustrated in figure 10.

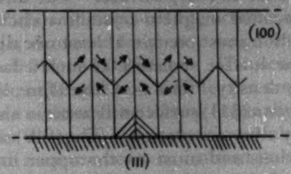


Figure 10. Idealized 90° wall structure in crystal with (111) faces.

The optimum domain size for such a structure may be calculated. The crystal faces (shaded) are again supposed to be (111) planes, and the surface structures are assumed to be like the *a*-domains in figure 8. Then the free energy per unit area of the (111) surface is

$$F=2(\tfrac{1}{8}K\gamma\delta)^{1/2}+\tfrac{1}{2}\gamma\frac{l}{\delta}+\sqrt{2}\gamma. \quad (\gamma\gamma_{180}\simeq=2\gamma_{90}).$$

δ is the main domain width and l is the perpendicular distance between the (111) faces. $(\partial F/\partial\delta)_l=0$ gives

$$\delta=(2\gamma K^{-1})^{1/3}\, l^{2/3}. \qquad \ldots\ldots(5)$$

δ is plotted against l in figure 9, curve B. It shows that the domain width with 90° walls is almost the same as that for the 180° wall structure over the whole range of crystal sizes.

It is interesting to compare the free energies of the two possible main domain structures, having 180° and 90° walls respectively. Expressions for these energies have been given above, and values corresponding to optimum domain sizes are plotted in figure 9 for crystals bounded, in both cases, by (111) faces. Over the whole range of crystal sizes the respective energies differ little from each other, though the 90° structure is slightly to be preferred in very small crystals. The fact that in most crystals in the demagnetized state the domain structure is observed to be an intimate mixture of both 90° and 180° walls is, therefore, not surprising.

§ 4. CONCLUSIONS

Optimum domain sizes in crystals with various surface orientations have been calculated, special attention being paid to the (010) and (111) cases respectively. It is not yet clear to what extent a demagnetized crystal may achieve its *optimum* domain size since the various processes involved in changes of magnetic state are seldom reversible and domain nucleation is normally guided by crystal imperfections (Bates and Martin 1956). Previous measurements as well as those recorded here have shown, however, that domains in large single crystals of soft magnetic materials are within a factor of about 2 of the sizes predicted by calculation. The values for the domain wall energy used in most calculations for iron are between 1 and 2 erg cm^{-2} and are derived on mainly theoretical grounds (Kittel 1949) and serve as an order of magnitude estimate only, so that better agreement should not in any case be expected. The present results confirm the estimates for iron.

A domain structure may be constructed for a surface of arbitrary orientation along the lines of that illustrated in figure 8. It is not difficult, for example, to imagine intermediate stages between the structure of figure 8 for a (111) surface and that described above for an (0*kl*) surface. The direction in a (111) surface having the smallest associated magneto-crystalline energy is the [110]; a face having any other orientation must contain at least one direction having a smaller energy than this, and the surface structure at such a face would have a correspondingly smaller energy density. It is clear, therefore, that for crystals of a given length in the [010] direction a (111) surface will result in a smaller optimum domain size than a surface of any other orientation. The (010) surface on the other hand contains two easy directions and must set the upper limit in optimum domain sizes. The figures recorded here thus set the range within which the optimum domain size for any crystal falls (except, perhaps, for certain crystals having one dimension much longer or shorter than the others).

The above calculations were concerned with single crystals. There is some evidence, however, that surface structures similar to those observed in single crystals also form at the faces of polycrystallites. The range of domain sizes recorded here should, therefore, be useful for estimating domain sizes in polycrystalline samples of iron. Surface magnetostatic energies, in fact, would be less for a polycrystallite and the lower limit to domain size might well move even closer to the upper limit set by the simple (010) structures.

ACKNOWLEDGMENTS

My thanks are due to Professor L. F. Bates for his encouragement and interest in the work, and to the Electrical and Allied Industries Research Association for past support.

REFERENCES

BATES, L. F., and MARTIN, D. H., 1956, *Proc. Phys. Soc.* B, **69**, 145.
BATES, L. F., and MEE, C. D., 1952, *Proc. Phys. Soc.* A, **65**, 129.
BOZORTH, R. M., 1951, *J. Phys. Radium.*, **12**, 308.
KITTEL, C., 1949, *Rev. Mod. Phys.*, **21**, 541.
LANDAU, L., and LIFSHITZ, E., 1935, *Phys. Z. Sowjet.*, **8**, 153.
LAWTON, H., 1949, *Thesis*, University of Cambridge.
WILLIAMS, H. J., BOZORTH, R. M., and SHOCKLEY, W., 1949, *Phys. Rev.*, **75**, 155.

The Effect of Gas Bubbles on Sound Propagation in Water

By J. D. MACPHERSON

Communicated by R. W. B. Stephens ; MS. received 18th June 1956, and in final form 10th September 1956

Department of Physics, Imperial College, London

Abstract. The effect of thin uniform bubble screens on sound propagation in water has been investigated over a frequency range of 15–100 kc/s for bubbles with diameters between 0·016 and 0·050 cm. The density of the bubble screen was varied from 1 bubble in 10 cm^2 to 4 bubbles per cm^2. The results obtained confirm to a considerable accuracy the theory originally proposed by Foldy and also provide values of the damping constants of gas bubbles which are consistent with those obtained from measurements on isolated bubbles.

---

§ 1. Introduction

In 1944 Foldy propounded a general theory for the propagation of sound in a liquid containing gas bubbles, which several workers have sought to verify experimentally (Corstensen and Foldy 1947, Laird and Kendig 1952, Meyer and Skudrzyk 1953, Fox, Curley and Larson 1955). Apart from Fox, Curley and Larson in 1955, the results obtained provide only an 'order of magnitude' agreement with the theory, and even their measurements are somewhat restricted by the same troubles as experienced by the previous workers. These difficulties have mainly been associated with the wide size distribution of the bubbles present and their 'non-uniform' distribution in space. These two factors made interpretation of the results difficult and caused a very considerable fluctuation with time of the measured quantities. Laird and Kendig (1952) for example, obtained values for the attenuation which varied by as much as 40 dB in a few seconds. A further inconvenience arose from the large attenuations produced by the bubbles, which necessitated either the placing of the two transducers only a few centimetres apart or the use of 'screens' of bubbles fairly thin but in practice rather poorly defined. The first method has the usual disadvantages associated with standing waves and diffraction patterns near the transducer radiating surface, while the second suffers from the obvious disadvantage that the boundaries, and thus the thickness of the screen, are uncertain.

The work described in this paper was initiated with the idea of obtaining a controlled system of bubbles on which accurate and repeatable measurements could be made. To achieve this objective it was necessary to produce a large number of bubbles of uniform size and to maintain a uniform distribution of these bubbles in space.

§ 2. Bubble Production

After several unsuccessful experiments, the method finally adopted for producing the bubbles was by electrolysis. This was carried out by applying a short voltage pulse to a series of very small similar electrodes. The electrode assembly consisted of a thick-walled glass tube about 25 cm long with short lengths of thermally pure platinum wire sealed into the glass wall along its length at regular intervals of about $\frac{1}{2}$ cm (along a generator of the cylinder). The outer surface of the tube was then ground down so as to leave a row of small disc-shaped electrodes exposed to the liquid. The wires were connected inside the tube by a mercury column.

When the voltage pulse is applied to this assembly, there is formed on each electrode a small bubble which is sufficiently buoyant to release itself immediately after it is formed. Thus by repeating the pulse at a certain rate, a two-dimensional lattice of slowly rising bubbles is formed.

This uniform thin bubble screen has the advantage that its boundaries are clearly defined and the attenuations produced by it are constant with time and of a suitable magnitude to measure. Thus the screen can be placed at a large distance from the transducers, so avoiding the possibility of the transducer radiation impedance being affected by the presence of the bubbles. At the same time, the uncertainty as to the form of the sound pressure field, caused by diffraction patterns near the radiating surface of the transmitter, is removed.

The range of bubble sizes which can be produced with a given electrode is determined mainly by its exposed surface area. The upper size limit is reached when the bubble becomes so large that part of it breaks away from the electrode before the end of the electrical pulse, the lower limit when the volume of gas liberated has insufficient buoyancy to release itself from the electrode. In the assembly used, the platinum wire had a diameter of 0·01 cm, which was the smallest practicable wire size, and bubbles with diameters of 0·016–0·050 cm could be produced by suitable adjustment of the voltage and its time of application.

Glass and platinum were chosen as the best combination for two reasons. Firstly, platinum seals into glass and is one of the metals most resistant to corrosion. Secondly, glass is 'wetted' by water, thus confining the gas to the electrode and preventing it from adhering to the surrounding glass. Different plastics were tried instead of glass but their large angles of contact with water made them unsuitable.

The range of lattice spacings obtained with such an electrode extended from about 1 bubble in 10 cm^2 up to approximately 4 bubbles per cm^2. The maximum density was limited firstly by the smallest horizontal spacing obtainable when the electrode assembly was made and secondly by the increasing danger of collision between two vertically adjacent bubbles as the vertical spacing is decreased.

The information concerning the distribution in spacing and size was obtained by means of a photograph given a finite exposure time. In this each rising bubble appears as a short trace, and from the length of the traces and the spacing between them it is possible to estimate the relative positions and sizes of the bubbles. In practice a standard deviation in radius of about 7% was achieved. The direct photographic method and the measurement of rate of rise were both found to be less accurate for determining the absolute bubble size than the acoustic method. The latter is dependent on the relation between size and resonant frequency which has been satisfactorily verified by several other workers (cf. Exner and Hampe 1953).

§ 3. THEORY

In the theory developed by Foldy it is shown that the part of the liquid containing the bubbles can usually be treated as a homogeneous medium with a complex propagation constant. In certain cases it is also necessary to consider the effect of the presence of a scattered or incoherent wave system. However, this is generally small and can in most practical cases be ignored.

In Corstensen and Foldy's paper it is shown that if a plane sound wave of the form

$$p = A_1 \exp\{jk(x\cos\theta + y\sin\theta) + j\omega t\}$$

is incident at an angle θ on a bubble screen of thickness s, then the transmitted and reflected amplitudes A_2 and B_1 respectively are given by

$$A_2 = A_1 \frac{\exp(jk_0 s\cos\theta)}{\cos(ks\cos\phi) + \frac{1}{2}j[k_0\cos\theta/k\cos\phi + k\cos\phi/k_0\cos\theta]\sin(ks\cos\phi)} \qquad \ldots\ldots(1)$$

$$B_1 = A_1 \frac{\frac{1}{2}j[k_0\cos\theta/k\cos\phi - k\cos\phi/k_0\cos\theta]\sin(ks\cos\phi)}{\cos(ks\cos\phi) + \frac{1}{2}j[k_0\cos\theta/k\cos\phi + k\cos\phi/k_0\cos\theta]\sin(ks\cos\phi)} \qquad \ldots\ldots(2)$$

where k, the complex propagation constant, is given by

$$k^2 = k_0^2 + 4\pi G \qquad \ldots\ldots(3)$$

and

$$G = \int_0^\infty g(R)n(r, R)dR \qquad \ldots\ldots(4)$$

where

$$g(R) = \frac{R}{\omega_r^2/\omega^2 - 1 + j\delta} \qquad \ldots\ldots(5)$$

R being the bubble radius, ω_r its resonant angular frequency and δ its damping constant. r is a radius vector from some arbitrary origin and $n(r, R)$ is the distribution function of bubble size and position. In a two-dimensional uniform rectangular lattice, the average number of bubbles per unit volume $n(r, R)dR$ is given by $1/abs$ where a and b are the lattice spacings.

For thin bubble screens of this type the thickness of the screen is very small, i.e. $s \sim R \ll \lambda$, where λ is the wavelength, so that certain approximations can be made: $\exp(jk_0 s\cos\theta) \simeq 1$, $\cos(ks\cos\phi) \simeq 1$, and $\sin(ks\cos\phi) \simeq ks\cos\phi$; also

$$\left[\frac{k_0\cos\theta}{k\cos\phi} + \frac{k\cos\phi}{k_0\cos\theta}\right] ks\cos\phi \simeq (k^2 - k_0^2)\frac{s}{k_0\cos\theta} + 2k_0 s\cos\theta \simeq (k^2 - k_0^2)\frac{s}{k_0\cos\theta} \qquad \ldots\ldots(6)$$

and

$$\left[\frac{k_0\cos\theta}{k\cos\phi} - \frac{k\cos\phi}{k_0\cos\theta}\right] ks\cos\phi = -(k^2 - k_0^2)\frac{s}{k_0\cos\theta}. \qquad \ldots\ldots(7)$$

With these results we find

$$\frac{A_1}{A_2} = 1 + j\frac{\lambda}{ab\cos\theta}g(R) \qquad \ldots\ldots(8)$$

$$\frac{A_1}{B_1} = j\frac{ab}{\lambda}\cos\theta\frac{1}{g(R)} - 1 \qquad \ldots\ldots(9)$$

and putting $\Delta=\omega_r^2/\omega^2-1$ and $K=\lambda R/ab$ the intensities are given by

$$\left|\frac{A_1}{A_2}\right|^2=1+\frac{K\sec\theta}{\Delta^2+\delta^2}[K\sec\theta+2\delta] \qquad \ldots\ldots(10)$$

$$\left|\frac{A_1}{B_1}\right|^2=1+\frac{2\delta}{K\sec\theta}+\frac{\Delta^2+\delta^2}{K^2\sec^2\theta}. \qquad \ldots\ldots(11)$$

When the bubbles are widely spaced so that $K\sec\theta\ll 2\delta$ equation (10) reduces to the result obtained by considering each bubble as an independent scattering element whose cross section is given by $4\pi\,|g(R)|^2$.

The phase difference ψ_T between A_1 and A_2 can be obtained from equation (8) and is given by

$$\tan\psi_T=\frac{\Delta}{\delta+(\Delta^2+\delta^2)/K\sec\theta}. \qquad \ldots\ldots(12)$$

Similarly, from equation (9), ψ_R the phase difference between A_1 and B_1 is given by

$$\tan\psi_R=-\frac{\Delta}{\delta+K\sec\theta}. \qquad \ldots\ldots(13)$$

These expressions are used to calculate the theoretical transmission through and reflection from the bubble screens.

To justify the omission of the incoherent component its magnitude should also be evaluated. Thus, following Foldy's method, the intensity of waves scattered once before reaching the hydrophone is given by

$$\frac{Q_1}{|A_1|^2}=\frac{R^2N\cos^2\theta}{d^2(\Delta^2+\delta^2)} \qquad \ldots\ldots(14)$$

where N is the total number of bubbles irradiated by the transmitter. On substituting typical values into this formula, it can be readily seen that for the cases under consideration, this term can be safely neglected.

§ 4. APPARATUS AND METHODS

The acoustic measurements on the bubble screen were carried out in a large unlined water tank 5 ft × 4 ft × 4 ft for which a pulse technique was necessary. A pulse (about 1 msec long) of sinusoidal oscillations was used for this purpose and the necessity of having at least about 15 complete cycles in the pulse enforced a lower frequency limit of 15 kc/s.

The apparatus was designed so that measurements could be made of the attenuation, phase shift and change in shape of the pulse as it was transmitted through the tank. A block diagram of this apparatus is shown in figure 1. The pulse unit supplied pulsed oscillations of the required frequency which were amplified and transmitted from an A.D.P. transducer. The 'pulse' was picked up by the hydrophone, a smaller A.D.P. transducer, amplified and displayed on a cathode-ray oscillograph. Its phase and amplitude relative to the input pulse were determined by making a direct observation when no bubbles were present in the tank. The pulse reflected from the bubble screen was picked up by another small A.D.P. transducer which had previously been calibrated with respect to the first-mentioned hydrophone.

§ 5. RESULTS

The smallest bubble size which could be reliably produced had a resonant frequency of approximately 30 kc/s. Since it was desired to cover as large a range of frequencies as possible, below as well as above the resonant frequency of the bubble screen, these smallest bubbles were used for the majority of the measurements. However, screens resonant anywhere in the range 15–40 kc/s were in fact produced and similar results obtained. The spatial distribution was varied between 1 and 40 bubbles in 10 cm^3.

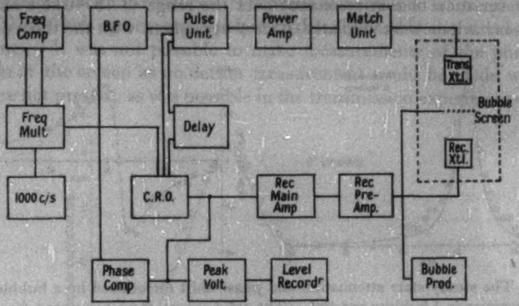


Figure 1.

The transmission measurements were made at angles of incidence of 0° and 45°. The bubble screen was placed directly behind a sheet 1 metre square of closed-cell rubber with a hole 15 cm square in its centre. This minimized the effects of diffraction around the edges of the bubble screen, as the attenuation produced by the rubber sheet was always large compared with the attenutation produced by the bubble screen.

In transmission, a bubble screen behaves as a mechanical band rejection filter. Thus, if a square pulse, figure 2 (for figures 2–6 see Plate), is applied to such a filter, then pulses of the form shown in figures 3 or 4 are produced (these photographs were obtained from electrical analogue circuits), provided that the signal frequency of the pulse is at or near the resonant frequency of the filter. From the exponential decay it is theoretically possible to calculate the damping factor or bandwidth of the system. Remote from resonance the effect of the filter is small and the output pulse will have approximately the same shape as the input pulse.

In practice these results were not fully realized. A typical pulse approximately resonant is shown in figure 5 and although it has the general form predicted it is subject to certain modifications. These are caused mainly by two effects: firstly, the mechanical pulse produced by the transmitter (figure 6) is not nearly as rectangular or free from spurious oscillations as the electrical pulse (figure 2), and secondly, the bubble screen is not exactly analogous to a simple filter, as it actually consists of a number of coupled resonant systems with a narrow spread of resonant frequencies. The shape of the transmitted pulse can not, therefore, be used for quantitative measurements.

Figure 7 shows the results obtained when the steady state phase shift and attenuation produced by a bubble screen of approximately 4 bubbles per cm^2 were measured. As can be seen, the agreement obtained with the theoretically predicted curves is good. The slight spread of the experimental points outside the theoretical curves can be ascribed to the small non-uniformity of the bubbles in the screen. Similar results were obtained with other spacings within the range previously mentioned. The measured value of the damping factor δ at 30 kc/s as obtained from equation (10) was found to be $0{\cdot}080 \pm 0{\cdot}003$ for oxygen bubbles. This agrees closely with the value predicted theoretically (0·079). The measured frequency variation of δ at resonance over the range of 15–40 kc/s and the off-resonance behaviour were similarly in close agreement with the theory.

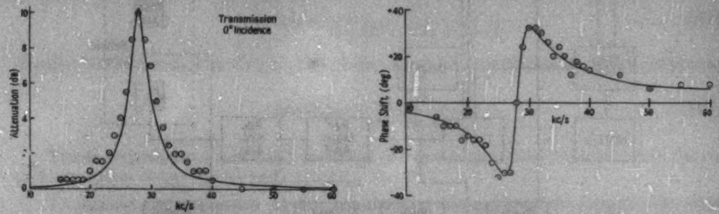


Figure 7. The steady state attenuation and phase shift introduced by a bubble screen of approximately 4 bubbles per cm^2. The solid curves were computed from equation (10) and (12) and the points are experimental observations.

Figure 8 shows the result of placing two screens of slightly different bubble sizes one behind the other, separated by a distance (~4 cm) which conforms to the condition for no interaction or wide spacing as obtained from equation (10). As can be seen the condition for no interaction is upheld in as much as the attenuations and phase shifts add algebraically.

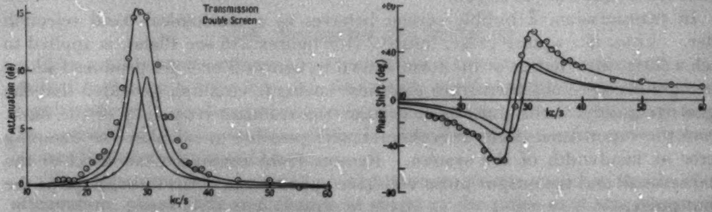


Figure 8. In this diagram the two smaller solid curves represent the attentuations and phase shifts introduced by the two screens individually and the larger solid curve is the algebraic sum of these two. As before the curves are computed from the theory and the points are experimental observations.

The reflection measurements were again made at 0° and 45° but without the rubber sheet. In reflection the bubble screen behaves as a band pass filter and, as previously for a square input pulse (figure 2) at or near the resonant frequency of the filter, pulses as shown in figures 9 and 10 (Plate) respectively are obtained. As before, away from resonance the change in pulse shape is small.

In practice a good approximation to this pulse shape is obtained (figure 11, Plate) as the extraneous oscillations produced by the transmitter are almost entirely transmitted through the screen. The value of δ obtained from the shape of the reflected pulse at 30 kc/s is $0{\cdot}09 \pm 0{\cdot}01$ and thus agrees fairly well with that obtained from the steady-state transmission and reflection measurements, the main source of error being that the decay is not exactly exponential because of the spread of bubble sizes.

The steady state reflection coefficient measurements, made on similar screens to those used for the transmission experiments, are shown in figure 12 and can be seen to agree well with the theory. The value of δ obtained from equation (11) $0{\cdot}080 \pm 0{\cdot}003$ at 30 kc/s agrees very closely with that obtained from the transmission measurements. It was not possible to make measurements of the phase shift on reflection at the screen as no datum measurement could be made when the bubbles were not present, as was possible in the transmission experiments.

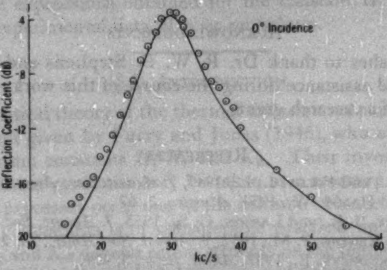


Figure 12. The steady state reflection coefficient of a bubble screen of approximately 4 bubbles per cm^2. The solid curves were computed from equations (11) and (13) and the points are experimental observations.

To provide experimental justification for calculating the theoretical curves from the coherent term only, some measurements were made to show the presence of the incoherent wave system. It was possible to show that at about 10 cm from a nearly resonant screen of 4 bubbles per cm^2 the measured intensity was of the order of magnitude given by the theory, namely about −12 dB. Away from resonance or further away from the bubble screen, it was not possible to detect this component.

§ 6. SUMMARY AND CONCLUSIONS

An experimental confirmation of Foldy's theory has been made within ±5% for the attenuation and reflection measurements and ±3° for phase shift measurements. The improved accuracy of the experimental results has been made possible by the virtual elimination of the fluctuations with time of the bubble sizes and of the spatial distributions. This enabled steady readings of attenuation, reflection coefficient and phase shift to be obtained. The range of bubble sizes used was from 0·016 to 0·050 cm in diameter (resonant from 40 to 15 kc/s) and the effective bubble densities from 1 bubble in 30 cm^3 to 8 bubbles per cm^3 as compared with the very restricted density ranges previously used, which were almost all in the region of 1 bubble per cm^3. With the bubble screens used, the resonance effect could be clearly observed and also accurate measurements could be obtained on non-resonant bubble systems over a wide range of frequencies, since the effect of these non-resonant bubbles was not masked by the presence of other resonant bubbles. The measurements made on the steady-state reflection coefficient, attenuation and phase shift permitted the calculation of the damping constant δ. This agreed well with the theoretically predicted value over the range of resonant frequencies covered and also, within the experimental limits, followed the off-resonance behaviour. The present results are therefore in agreement with those obtained by Exner and Hampe (1953) on single bubbles. The transient effects when a pulse was propagated through and reflected from a bubble screen were investigated and the expected deformations of the pulse shape observed and recorded. The value of the damping constant obtained from the exponential decay was in fair agreement with the value obtained from the steady state observations. Finally, it was shown that the incoherent or scattered component could in most practical cases be ignored.

ACKNOWLEDGMENTS

The author wishes to thank Dr. R. W. B. Stephens and Dr. G. G. Parfitt for their advice and assistance during the course of this work, and the Admiralty for the provision of a research grant.

REFERENCES

CORSTENSEN, F. L., and FOLDY, L. L., 1947, *J. Acoust. Soc. Amer.*, **19**, 481.
EXNER, M. L., and HAMPE, W., 1953, *Acustica*, **3**, 67.
FOX, F. E., CURLEY, S. R., and LARSON, G. S., 1955, *J. Acoust. Soc. Amer.*, **27**, 534.
LAIRD, D. T., and KENDIG, P. M., 1952, *J. Acoust. Soc. Amer.*, **24**, 29.
MEYER, E., and SKUDRZYK, E., 1953, *Akust. Beih.*, **3**, 434.

Clusius and Dickel Column Constants for a Gas consisting of Hard Spherical Molecules

By R. C. SRIVASTAVA
Department of Physics, University of Lucknow, India

Communicated by B. N. Srivastava; MS. received 1st May 1956, and in final form 4th September 1956

Abstract. The theory of the Clusius and Dickel hot wire thermal diffusion column as developed by Furry and Jones has been extended to take into account the temperature dependence of the gas coefficients for a gas whose molecules consist of hard rigid spheres. Tables for the shape factors h, k_c and k_d have been constructed by the method of numerical integration. It is found that the column constants H, K_c and K_d depend strongly on the law of molecular interaction for the gas. The expressions obtained for the constant H are found to agree closely with the experimental data of Nier on helium.

---

§ 1. Introduction

The mathematical theory of the thermal diffusion column for the cylindrical case was first given by Furry and Jones (1946), who obtained expressions for the column constants H, K_c and K_d. Their investigations show that certain ratios of these constants may be taken as a measure of the effectiveness of the separation processes occurring in the column. Hence a correct knowledge of values of these constants is of considerable practical importance. It can be shown that H, K_c and K_d depend rather strongly on the temperature dependences of the gas coefficients. Furry and Jones assume for the thermal conductivity, viscosity and diffusion constants of a gas a temperature dependence pertaining to the Maxwellian molecules for which $n=1$, n being related to the force constant ν by the relation $\nu=(2n+3)/(2n-1)$. But as the experimental values of n for nearly all the gases lie between $\frac{1}{2}$ and 1, calculations for the other limiting case when $n=\frac{1}{2}$ (when molecules are taken to be rigid spheres) are very desirable.

The present paper is devoted to the task of determining values of H, K_c and K_d for the extreme cylindrical case, when the gas coefficients have a temperature dependence pertaining to rigid spherical molecules ($n=\frac{1}{2}$). The variation of the thermal diffusion constant α, with temperature in the form $\alpha=A-(B'/T)$, where A and B' are constants for any gas and T is the absolute temperature, is also taken into account.

§ 2. Theory of the Extreme Cylindrical Case

Throughout this paper the terminology of Furry and Jones (to be referred to as FJ) will be used and their equations (106) to (126) will be utilized but, in order to avoid repetition, these equations are not given here.

We assume that the following quantities are independent of temperature: $\eta/T^{1/2}$, $\lambda/T^{1/2}$, $\rho D/T^{1/2}$; η/λ, $\rho D/\lambda$. Here ρ, D, λ and η are respectively the density, diffusion constant, thermal conductivity and coefficient of viscosity.

The heat flow by conduction per unit length of the annular space between cylinders of radii r and $r+dr$ is given by

$$2\pi Q=2\pi r\lambda(-dT/dr)\,\text{cal cm}^{-1}\,\text{sec}^{-1}.$$

This yields on integration

$$(3/2)\log r = -(\lambda/QT^{1/2})T^{3/2} + \text{constant}. \qquad \ldots\ldots(1)$$

We now introduce the dimensionless variable t and the dimensionless parameters t_1 and t_2 and a length r_0 given by the following equations:

$$t = (\lambda/QT^{1/2})^{2/3}T; \qquad t_i = (\lambda/QT^{1/2})^{2/3}T_i \quad (i = 1, 2) \qquad \ldots\ldots(2)$$

$$r^{3/2} = r_0^{3/2}\exp(-t^{3/2}). \qquad \ldots\ldots(3)$$

We also introduce a new master function $\gamma(t)$ which differs from $G(T)$ by a constant factor given by

$$G(T) = -\left(\frac{QT^{1/2}}{\lambda}\right)^{4/3}(\lambda^4\rho gT/\eta D)r_0^4\gamma(t). \qquad \ldots\ldots(4)$$

The differential equation

$$\frac{d}{dT}\frac{1}{\lambda r^2}\frac{d}{dT}\frac{\eta}{\lambda}\frac{d}{dT}\frac{1}{\lambda\rho r^2}\frac{d}{dT}\left(\frac{\rho D}{\lambda}G(T)\right) = g\frac{d\rho}{dT},$$

of FJ now takes the form

$$\frac{d}{dt}t^{-1/2}\exp(4t^{3/2}/3)\frac{d^2}{dt^2}t^{1/2}\exp(4t^{3/2}/3)\frac{d\gamma}{dt} = t^{-2} \qquad \ldots\ldots(5)$$

with the boundary conditions

$$\gamma(t_1) = \gamma(t_2) = \gamma'(t_1) = \gamma'(t_2) = 0. \qquad \ldots\ldots(6)$$

The constants H, K_c and K_d will be given by equations (9) to (14) of FJ but instead of the shape factors h, k_c and k_d given by

$$h = 6\,!\,t_1^3\exp(2t_1^2)\int_{t_1}^{t_2}(\alpha\gamma/t)\,dt, \qquad k_c = 9\,!\,t_1^7\exp(4t_1^2)\int_{t_1}^{t_2}\gamma^2\,dt$$

and

$$k_d = t_1^{-1}\exp(t_1^2)\int_{t_1}^{t_2}t^2\exp(-t^2)\,dt,$$

we shall now have

$$h = 6\,!\,t_1^{5/2}\exp(8t_1^{3/2}/3)\int_{t_1}^{t_2}(\alpha\gamma/t)\,dt, \qquad \ldots\ldots(7)$$

$$k_c = 9\,!\,t_1^{11/2}\exp(16t_1^{3/2}/3)\int_{t_1}^{t_2}\gamma^2\,dt, \qquad \ldots\ldots(8)$$

and

$$k_d = \tfrac{1}{2} - \tfrac{1}{2}(t_2/t_1)^{1/2}\exp\{4(t_1^{3/2} - t_2^{3/2})/3\} + (f_{12}/4t_1^{1/2})\exp(4t_1^{3/2}/3). \qquad \ldots\ldots(9)$$

§ 3. EXPRESSION FOR THE FUNCTION $\gamma(t)$

It can easily be shown that the solution of the equation (5) which satisfies equation (6) will again be given by the equations (27) to (29) of FJ and all their equations (59) to (73) will be valid for our case as well. However, the various functions involved in these expressions will now be given by the following equations:

$$f = \int^{t} x^{-1/2}\exp(-4x^{3/2}/3)\,dx, \qquad f_i = \int^{t_i} x^{-1/2}\exp(-4x^{3/2}/3)\,dx, \quad i = 1, 2.$$

$$f_{t\infty} = \int_{t}^{\infty} x^{-1/2}\exp(-4x^{3/2}/3)\,dx; \qquad f_{12} = \int_{t_1}^{t_2} x^{-1/2}\exp(-4x^{3/2}/3)\,dx. \qquad \ldots\ldots(10)$$

$$g = \int^{t} x^{-1/2}\exp(-8x^{3/2}/3)\,dx, \qquad g_i = \int^{t_i} x^{-1/2}\exp(-8x^{3/2}/3)\,dx,$$

$$g_{t\infty} = \int_{t}^{\infty} x^{-1/2}\exp(-8x^{3/2}/3)\,dx. \qquad \ldots\ldots(11)$$

We shall write e for $\exp(-4t^{3/2}/3)$ and e_i for $\exp(-4t_i^{3/2}/3)$, $i=1, 2$.

If the temperature variation of thermal diffusion constant, α in the form $\alpha = A - B'/T$ (Srivastava 1955) is taken into account, the expressions for k_c and k_d will remain unchanged as they do not contain α, but for h, instead of equations of FJ, viz.

$$h \simeq 6!\, t_1^{3} \exp(2t_1^{2}) \frac{h_\infty - (h_1''/t_2)}{1-(\delta''/t_2)}, \qquad \ldots\ldots(12)$$

where

$$\delta'' = (e_1/f_{1\infty}) - t_1,\ h_\infty = \int_{t_1}^{\infty} (\alpha\gamma_\infty/t)\,dt \qquad \text{and} \qquad h'' = \int_{t_1}^{\infty} (\alpha\gamma''/t)\,dt$$

we now have

$$h \simeq 6!\, t_1^{5/2} \exp(8t_1^{4/3}/3) \frac{h_1 - (h_1''/t_2)}{1-(\delta''/t_2)}, \qquad \ldots\ldots(13)$$

with

$$h_1 = A\int_{t_1}^{\infty} (\gamma_\infty/t)\,dt - (B't_1/T_1)\int_{t_1}^{\infty} (\gamma_\infty/t^2)\,dt = Ah_\infty - (B't_1/T_1)h_{\infty B}, \qquad \ldots\ldots(14)$$

where $$h_\infty = \int_{t_1}^{\infty} (\gamma_\infty/t)\,dt \qquad \text{and} \qquad h_{\infty B} = \int_{t_1}^{\infty} (\gamma_\infty/t^2)\,dt, \qquad \ldots\ldots(15)$$

and

$$h_1'' = A\int_{t_1}^{\infty} (\gamma''/t)\,dt - (B't_1/T_1)\int_{t_1}^{\infty} (\gamma''/t^2)\,dt = Ah'' - (B't_1/T_1)h_B''$$

where $$h'' = \int_{t_1}^{\infty} (\gamma''/t)\,dt \quad \text{and} \quad h_B'' = \int_{t_1}^{\infty} (\gamma''/t^2)\,dt. \qquad \ldots\ldots(16)$$

§ 4. NUMERICAL DATA FOR γ_∞ AND γ'' AND THE RESULTS

In order to calculate the values of γ_∞ and γ'' a correct knowledge of all the functions involved therein viz. e, f, g etc. is essential. Since tables are not available for these functions numerical calculations, correct to the seventh place of decimals were performed.

The values of γ_∞ and γ'' for certain fixed values of t_1 and several values of t are given in table 1. These values may be used for numerical calculations for any specified temperature dependence of α. From a knowledge of γ_∞ and γ'' the coefficients h, k_c and k_d were calculated by numerical integration. These are given in table 2.

Using the results of table 2, we can compute from our equation (9) and FJ's equations (70) and (71) values of k_d, h and k_c, respectively, for a chosen set of values of $r_1/r_2 \gg 1$ and for certain values of T_2/T_1 which are determined by available values of t_1. By graphical interpolation we can then obtain h, k_c and k_d for other values of T_2/T_1 (tables 3, 4 and 5). For the sake of comparison the corresponding values obtained by FJ have also been put side by side in these tables.

A comparison between the values of h, k_c and k_d for $n=\frac{1}{2}$ and for $n=1$ shows that the former are considerably lower. This indicates that in the extreme cylindrical case of a thermal diffusion column H, K_c and K_d depend very strongly on the value of n. Hence for a correct evaluation of these constants the appropriate value of n for the gas under investigation must be taken into account. For values of n between $\frac{1}{2}$ and 1, h, k_c and k_d can be obtained by graphical interpolation.

Table 1. Values of γ_∞ and γ''

t	$t_1=0\cdot5$		$t_1=0\cdot6$		$t_1=0\cdot7$		$t_1=0\cdot8$	
	$\gamma_\infty\times10^6$	$\gamma''\times10^6$	$\gamma_\infty\times10^6$	$\gamma''\times10^6$	$\gamma_\infty\times10^6$	$\gamma''\times10^6$	$\gamma_\infty\times10^6$	$\gamma''\times10^6$
0·5	0	0						
0·55	9	15						
0·60	29	49	0	0				
0·65	60	101	8	15				
0·70	81	137	20	35	0	0		
0·75	104	178	37	64	3	6		
0·80	122	212	53	93	12	22	0	0
0·85	133	244	63	125	17	33	2·7	5·8
0·90	147	258	80	142	32	59	7·6	14
1·00	155	278	93	171	48	90	20	38
1·05	150	273	95	175	51	98	23	46
1·10	144	266	93	174	53	103	26	55
1·15	137	257	91	173	52	106	29	58
1·20	125	240	83	163	51	101	26	56
1·30	107	212	73	149	47	98	25	55
1·40	85	176	60	126	38	84	22	52
1·50	64	141	44	101	28	67	15	41
1·60	49	113	30	82	15	45	8	25
1·70	34	97	13	58	9	17	2·3	9
1·80	14	66	2·5	18				
1·90	1·7	21						

Table 2. Values of h, k_c, δ'', etc.

t_1	0·5	0·6	0·7	0·8
h_∞	237×10^{-6}	118×10^{-6}	$51\cdot3\times10^{-6}$	$22\cdot0\times10^{-6}$
$h_{\infty B}$	475×10^{-6}	196×10^{-6}	$73\cdot4\times10^{-6}$	$27\cdot2\times10^{-6}$
h''	456×10^{-6}	227×10^{-6}	108×10^{-6}	$48\cdot5\times10^{-6}$
h''_B	903×10^{-6}	390×10^{-6}	155×10^{-6}	$60\cdot6\times10^{-6}$
k_∞	132×10^{-6}	$46\cdot6\times10^{-6}$	$14\cdot6\times10^{-6}$	$3\cdot71\times10^{-6}$
k_1''	238×10^{-6}	$92\cdot8\times10^{-6}$	$29\cdot3\times10^{-6}$	$7\cdot98\times10^{-6}$
k_2''	454×10^{-6}	182×10^{-6}	$60\cdot1\times10^{-6}$	$18\cdot5\times10^{-6}$
δ''	1·2586	1·3467	1·4341	1·5212

Table 3. Values of h with α constant

r_1/r_2	15		25		40		60		100	
T_2/T_1	$n=\frac{1}{2}$	$n=1$ (FJ)	$n=\frac{1}{2}$	$n=1$ (FJ)	$n=\frac{1}{2}$	$n=1$ (FJ)	$n=\frac{1}{2}$	$n=1$ (FJ)	$n=\frac{1}{2}$	$n=1$ (FJ)
3	0·015	0·091	0·021	0·098						
4	0·036	0·092	0·037	0·103	0·035	0·109	0·033	0·113	0·032	0·116
5	0·045	0·075	0·052	0·092	0·050	0·103	0·047	0·108	0·045	0·114
6	0·040	—	0·045	—	0·055	0·085	0·061	0·093	0·058	0·102
6·5	—	—	—	—	0·054	—	0·055	—	0·062	—

Table 4. Values of k_c

r_1/r_2	15		25		40		60		100	
T_2/T_1	$n=\frac{1}{2}$	$n=1$ (FJ)	$n=\frac{1}{2}$	$n=1$ (FJ)	$n=\frac{1}{2}$	$n=1$ (FJ)	$n=\frac{1}{2}$	$n=1$ (FJ)	$n=\frac{1}{2}$	$n=1$ (FJ)
3	0·00089	0·0095	0·00110	0·0130						
4	0·00066	0·0045	0·00083	0·0068	0·00099	0·0088	0·00120	0·0105	0·00121	0·0128
5	0·00043	0·0022	0·00059	0·0034	0·00076	0·0046	0·00085	0·0056	0·00097	0·0072
6	0·00020	—	0·00035	—	0·00050	0·0025	0·00061	0·0031	0·00076	0·0040
6·5	—	—			0·00038	—	0·00048	—	0·00054	—

Table 5. Values of k_d

r_1/r_2	15		25		40		60		100	
T_2/T_1	$n=\frac{1}{2}$	$n=1$ (FJ)	$n=\frac{1}{2}$	$n=1$ (FJ)	$n=\frac{1}{2}$	$n=1$ (FJ)	$n=\frac{1}{2}$	$n=1$ (FJ)	$n=\frac{1}{2}$	$n=1$ (FJ)
3	0·60	0·75	0·59	0·73	—	—	—	—	—	—
4	0·64	0·91	0·62	0·87	0·61	0·83	0·60	0·81	0·59	0·78
5	0·68	1·08	0·66	1·01	0·65	0·97	0·64	0·94	0·62	0·90
6	0·72	—	0·69	—	0·68	1·11	0·67	1·07	0·66	1·03
6·5	—	—	—	—	0·70	—	0·69	—	0·67	—

§ 5. COMPARISON WITH EXPERIMENTAL RESULTS

We now proceed to test the results obtained here by comparison with the experimental data on helium. We have selected helium since for this gas the value of n lies between 0·644 and 0·669 (Jones and Furry 1946) which is nearer to $\frac{1}{2}$. Moreover, sufficiently accurate data due to Macinteer, Aldrich and Nier (1948) for the extreme cylindrical case are available for helium. Due to the uncertainty of the asymmetry term K_p, a correct theoretical value of K $(=K_c+K_d+K_p)$, and hence also of $2A$, which is equal to H/K cannot be obtained. Therefore, only experimental and theoretical values of H are compared. Unfortunately, the experimental value of α is given for only one temperature and therefore we cannot account for the temperature variation of α. However, its contribution will be much smaller than that produced by the change in the value of n. The results are given in table 6.

Table 6. Values of H in standard litres per day

Pressure (atm)	H_{exp}	H_{theor} for		
		$n=\frac{1}{2}$	$n=\frac{2}{3}$	$n=1$
7·8	9·5	5·4	9·1	16·5
9·7	14·8	8·1	13·8	24·8

It is found that the agreement between the experimental and the theoretical values of H for $n=\frac{1}{2}$ for helium is not good, though definitely better than the corresponding values of H for $n=1$. However, if we find by interpolation the value of H for $n=\frac{2}{3}$, which is the correct value of n for helium, the agreement with the experimental results is fairly good. A still better agreement is expected if the variation of α with temperature is also taken into account.

This, in our opinion, provides the clue to the well-known experimental fact that the required length of the column is actually about 40% larger than that given by the theory of Furry and Jones. It seems fairly certain that for all practical purposes the correct column length could be obtained from a proper choice of n.

ACKNOWLEDGMENTS

The author is deeply indebted to Professor B. N. Srivastava, who suggested this problem, for his continued advice and interest in the work and to Professor P. N. Sharma, for valuable suggestions and constant encouragement. Thanks are also due to Mr. B. N. Joshi for his help in verifying the mathematical steps.

REFERENCES

FURRY, W. H., and JONES, R. C., 1946, *Phys. Rev.*, **69**, 459.
JONES, R. C., and FURRY, W. H., 1946, *Rev. Mod. Phys.*, **18**, 151.
MACINTEER, B. B., ALDRICH, L. T., and NIER, A. O., 1948, *Phys. Rev.*, **74**, 946.
SRIVASTAVA, R. C., 1955, *Proc. Phys. Soc.* A, **68**, 294.

The Adhesion Theory of Friction

By R. T. SPURR and T. P. NEWCOMB
Ferodo Limited, Chapel-en-le-Frith, Stockport, Cheshire

MS. received 31*st July* 1956

Abstract. It is shown that the static friction between a specimen of bituminous material and a cast iron disc is due to adhesion between specimen and disc, and that the friction is reduced when the normal load is removed because of elastic interaction between the surfaces.

---

§ 1. INTRODUCTION

TRUE contact between two surfaces occurs only at small areas dispersed over the nominal area of contact. It is generally considered that adhesion occurs between the surfaces at these contact areas and that the frictional resistance to sliding is due to the force required to overcome this adhesion. This theory has been criticized (e.g. Bikerman 1941) on the grounds that if the adhesion is large enough to be the cause of frictional resistance it should be experimentally measurable when separating the surfaces normally, but in general no such adhesion has been observed. This criticism has been ingeniously met by Bowden and Tabor (1939) who suggested that when the normal load is removed elastic recovery of the surfaces breaks the junctions formed at the localized contact areas and so adhesion is not detectable. Bowden and Rowe (1956) have shown that adhesion can occur between metals if they have been heated above their annealing temperatures in order to reduce the elastic stresses at the junctions.

The bituminous material 'Mexphalte', (from the Shell Petroleum Co. Ltd.) adheres to a cast iron surface and in the course of an investigation into the relationship between the adhesion and friction of 'Mexphalte', some observations have been made which support the adhesion theory and show the importance of elastic recovery.

§ 2. EXPERIMENTAL DETAILS

Nominally flat specimens of 'Mexphalte' (about 1 in. square) were used. The specimens were fixed in a hinged mount at the end of a torque arm which could rotate freely about the same axis as a polished cast iron disc (35 cm in diameter) driven by a 3 h.p. electric motor. The disc could be heated to various temperatures by a length of electrical heating strip. Before making any measurement the specimen was 'run in' under load against the rotating disc until it had a smooth polished appearance. The disc was then clamped stationary, the specimen was loaded against it for a set period of time and the friction measured by pulling at the torque arm with a spring gauge until the specimen slipped. Similarly, to measure the adhesion a vertical force was applied through a spring gauge, the counterweighted hinged mount permitting the specimen to swing up from the plate when the adhesion was overcome. Loads of from 150 to 2000 grammes were used. The measured values of the friction and the adhesion were dependent upon the time the surfaces had been in contact. The static coefficient of friction measured after the surfaces had been loaded together for 30 seconds was called μ_{on}. Other measurements were made in which the normal load was removed after 30 seconds and then 30 seconds later the tangential force necessary to shear through the junction formed between specimen and disc measured. From this force

$$\mu_{\mathrm{off}} = \text{(tangential force)/(normal load)}$$

was obtained. The coefficient of adhesion

$$\sigma_{\mathrm{off}} = \text{(normal load)/(adhesion)}$$

was measured in the same way as μ_{off} except that a normal instead of a tangential force was applied. As there was some scatter in the results, a number of sets of measurements, each set consisting of adhesion and friction measurements using the one specimen, was made over the range of loads, several measurements being made at each load. Measurements in each set were made in random order.

§ 3. Results and Discussion

It was found that over the experimental range of loads μ_{on} was proportional to $L^{-0\cdot 60}$; μ_{off} to $L^{-0\cdot 70}$ and σ_{off} to $L^{-0\cdot 74}$, where L is the normal load. It appears that within the experimental error of the measurements $\mu_{\mathrm{off}} = K\sigma_{\mathrm{off}}$ with $K = 1\cdot 35$. In figure 1 the frictional force (normal load removed) is plotted against adhesion for the whole range of loads. In the σ_{off} measurements the tensile strength of the junctions was measured whereas in the μ_{off} measurements the shear strength was measured; apparently the tensile strength of a junction is a definite fraction of its shear strength.

μ_{on} was greater than μ_{off} at room temperature. To determine whether the difference could be due to the areas of contact increasing in size when the tangential load was applied to measure μ_{on} (McFarlane and Tabor 1950) the normal load was applied for 30 seconds, a tangential load just short of that necessary to bring about slip then applied to the specimen for 5 seconds, the tangential and normal loads removed and 30 seconds later the adhesion measured. If the

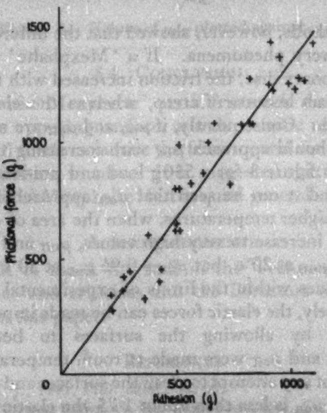


Figure 1. Frictional force (after removal of normal load) plotted against adhesion.

tangential load had caused an increase in contact area, the adhesion after the tangential load had been applied should be correspondingly greater than the value obtained when no such load had been used. There was a small increase but this was found to be due to the extra five seconds of normal loading.

μ_{on} could also be greater than μ_{off} because of elastic recovery reducing the true contact area when the normal load was removed to measure μ_{off}. To obtain an idea of the magnitude of the visco-elastic recovery in 'Mexphalte', $\frac{1}{8}$ in. hemispheres were loaded against a glass plate and the area of contact measured with a microscope. The load was applied for 120 seconds and then removed; a typical area–time curve is shown in figure 2, and it can be seen that the recovery is considerable. Further μ_{off} measurements were therefore made, but the interval between the removal of the normal load and the measurement of μ_{off} was varied between 1 and 120 seconds in the expectation that μ_{off} should approach μ_{on} as the interval was decreased. However, any variation in μ_{off} with variation in the time between removing the load and measuring the friction was smaller than the experimental error in the measurements. Presumably, the recovery in nominal flats is almost entirely elastic and consequently very rapid.

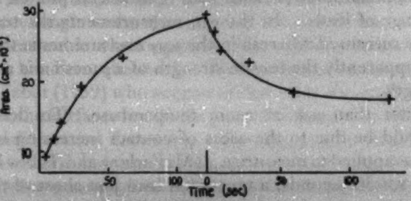


Figure 2. Creep and recovery curve of 'Mexphalte'. The normal load was removed after 120 seconds.

More indirect methods, however, showed that the difference between μ_{on} and μ_{off} was due to recovery phenomena. If a 'Mexphalte' specimen was left in contact with the stationary disc, the friction increased with the length of time the surfaces were in contact because of creep, whereas the elastic forces remained practically unchanged. Consequently, if μ_{on} and μ_{off} are measured for different pre-load times, μ_{off} should approach μ_{on} with increasing times. Values of μ_{on} and μ_{off} are shown in figure 3 for a 550 g load and pre-load times varying from 2 to 1000 seconds, and it can be seen that μ_{off} approaches μ_{on} with increasing time. Similarly, at higher temperatures, when the area of contact and therefore the adhesion and μ_{on} increase to very high values, μ_{off} and μ_{on} converge, e.g. at 550 g load $\mu_{\text{off}} = 0{\cdot}85\ \mu_{\text{on}}$ at 20°C, but $\mu_{\text{off}} = 0{\cdot}95\ \mu_{\text{on}}$ at 30°C, and at 35°C μ_{on} and μ_{off} had the same values within the limits of experimental error over the entire load range. Conversely, the elastic forces can be made important compared with the adhesive forces by allowing the surfaces to become contaminated. Measurements of μ_{on} and μ_{off} were made at room temperature at various times with 550 g load without any attempt to clean the surfaces and are shown in figure 4. It can be seen that if μ_{on} is less than about 1·75 the elastic recovery is sufficient to overcome the adhesion, i.e. $\mu_{\text{off}} = 0$.

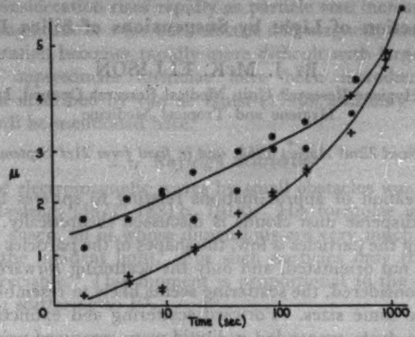


Figure 3. Values of μ_{on} (upper curve) and μ_{off} (lower curve) plotted against log (pre-load time).

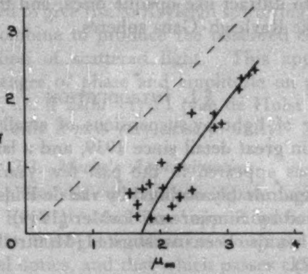


Figure 4. Values of μ_{off} plotted against μ_{on}.

§ 4. CONCLUSION

It appears that, with 'Mexphalte', static friction is due to adhesion between the specimen and the opposing surface, and that elastic recovery of the surfaces on removal of the load tends to reduce the friction, and can, if the adhesion is low enough, reduce the friction to zero. It is suggested that such behaviour is not limited to materials like 'Mexphalte' but is quite general.

ACKNOWLEDGMENTS

The authors wish to thank the staff of the Technical Division of Ferodo Limited for assistance and criticism and the Directors of Ferodo Limited for permission to publish this paper.

REFERENCES

BIKERMAN, J. J., 1941, *Phil. Mag.*, **32**, 69.
BOWDEN, F. P., and ROWE, G. W., 1956, *Proc. Roy. Soc.* A, **233**, 429.
BOWDEN, F. P., and TABOR, D., 1939, *Proc. Roy. Soc.* A, **169**, 391.
MCFARLANE, J. S., and TABOR, D., 1950, *Proc. Roy. Soc.* A, **202**, 244

Extinction of Light by Suspensions of Silica Dust†

By J. McK. ELLISON

Environmental Hygiene Research Unit, Medical Research Council, London School of Hygiene and Tropical Medicine

MS. received 22nd August 1956, *and in final form* 21*st September* 1956

Abstract. Application of approximations relating to spheres to the scattering of light by polydisperse dust clouds is discussed theoretically. If the relative refractive index of the particles is low, the shapes of the particles are not extreme, the particles are not orientated, and only the scattering forward through angles less than 30° is considered, the scattering seems likely to resemble that by clouds of spheres of the same sizes. Forward scattering and extinction by carefully fractionated silica dusts suspended in liquid were measured using an extinction method, and for particles larger than 2 microns showed good agreement with Mecke's approximation. For smaller particles absolute extinction measurements were not made, but relative values indicated the presence of aggregates. The aggregates seemed to diffract like opaque discs, and the unaggregated particles seemed to scatter like Rayleigh–Gans spheres.

§ 1. Introduction

The scattering of light by suspensions of small transparent spheres has been studied in great detail since 1939, and a large number of papers on this subject has appeared in the past few years. The general features of this scattering are admirably outlined by van de Hulst (1946, 1949) and recent advances are reviewed by Sinclair and LaMer (1949). Scattering by cylinders, discs, and ellipsoids has also been investigated (Montroll and Hart 1951, Montroll and Greenberg 1952).

Scattering by powdered material, however, has not received the same attention. This is not surprising, for the particles present are of all shapes and sizes, and the physical processes occurring cannot be defined in detail. Nevertheless, light scattering by dusts is of considerable interest. It is the purpose of this paper to see how closely the forward scattering and extinction approximate to those by spheres of the same size and relative refractive index.

§ 2. Light Scattering by Transparent Spheres

An exact solution of the equations giving the scattering of electromagnetic waves by spherical obstacles was first obtained by Mie (1908). This solution appears as a series of terms expressing the incident and scattered fields in terms of the radiation by electric and magnetic dipoles and multipoles. The variables appearing in the solution are (i) m, the refractive index of the sphere relative to that of the medium in which it is suspended; (ii) $x=2\pi r/\lambda$, where r = radius of sphere and λ = wavelength of incident light, measured in the suspending medium; and (iii) θ, the angle between the directions of propagation of incident and scattered beams. The effective number of multipole components which must be taken into consideration rises rapidly as particle size increases, and is of the same order of magnitude as x for both electric and magnetic multipoles. Consequently, computation becomes rapidly more difficult with large particles. As a result, various approximate formulae have been introduced. These are particularly well described by van de Hulst (1946), and only those relevant to dust particles will be mentioned here.

---

† This paper is based on part of a thesis accepted by the University of London for the Ph.D. degree.

2.1. *Rayleigh Scattering*

Scattering of electromagnetic waves by small obstacles was first successfully explained by Rayleigh (1871, 1881 b, 1918). His formulae apply to isotropic particles of any shape, all of whose dimensions are very much smaller than the wavelength of the incident light. For such particles only the electric dipole moment is appreciable. This moment is proportional to the volume V of the particle, and the scattering to V^2.

2.2. *The Rayleigh–Gans Approximation*

Many years later Rayleigh (1910), and independently Gans (1925), derived the Rayleigh–Gans approximation. This is applicable to larger particles of low refractive index relative to the medium in which they are suspended. Each volume element is assumed to give rise to Rayleigh scattering, and the amplitudes of the scattered waves combine to produce the observed effects. Interference determines the distribution of scattered light. This approximation can be employed only if the changes of phase and amplitude on passing through the particle can be neglected, i.e. if $2x(m-1) \ll 1$ (van de Hulst 1946). If $x > 5$ this condition can be satisfied only if m is very close to unity.

2.3. *Mecke's Approximation*

If $2x(m-1) \simeq 1$, the phase-shift of rays passing through the sphere cannot be neglected. Instead of Rayleigh's and Gans' approximation, however, Mecke's (1920) may be used. By it, the light incident on the sphere is treated according to the laws of geometrical optics, and that which passes close to it is treated by applying diffraction theory to a disc of the same radius. These two components recombine to give the observed scattering, but if $2x(m-1) \gg 1$ interference between them is unimportant and they may be treated independently (van de Hulst 1946, Bricard 1946). Then since each component involves a flux equal to that incident on the cross section of the sphere, the ratio of light scattered and absorbed by the particle to that incident on its geometrical cross section is 2. This ratio is called the extinction E. The apparent anomaly of an obstacle scattering more light than is incident on its geometrical cross section is a general feature of the scattering of waves by large obstacles, and is discussed in detail by Brillouin (1949).

§ 3. SCATTERING BY NON-SPHERICAL TRANSPARENT PARTICLES

In principle the methods of approximation outlined above are clearly applicable to particles of other shapes. The applicability of the Rayleigh formula to non-spherical particles has already been mentioned. Both of the other two approximations require some modification, however.

Each is derived from an integration—that in the Rayleigh–Gans case being over the volume, and that in the Kirchhoff diffraction of Mecke scattering being equivalent to one over the area of cross section. With non-spherical particles, if the integration is taken over the corresponding volume or area, and, in the Mecke case, if the corresponding geometrical component is then added, an analogous approximation results. This should be equally good if the particles are not extreme in shape. This method has been applied to the determination of the shapes of polymer molecules in solutions (see for example Doty and Steiner 1950) on the assumption that they scatter according to the Rayleigh–Gans approximation, and that they are unorientated. The shapes assumed are uniform and extreme, despite which the differences are not very large for small values of $z = x \sin\theta$. With dusts consisting mainly of small compact particles whose largest dimension d is such that $\pi d/\lambda \not\gg 1$, the scattering forward may be expected to resemble that by spheres of the same size.

For Kirchhoff diffraction the same is true. Vouk (1948) pointed out that the radial distribution of the light scattered by opaque screens was not sharply dependent on their shape, provided this was not extreme.

If the particles are transparent the geometrical scattering must be considered. Wide-angle scattering is not of great interest, since the diffraction effects are not readily interpreted and are indistinguishable from the geometrical. With spheres, geometrical scattering through small angles arises predominantly from rays which have either been externally reflected or have passed through the particle without suffering internal reflection. Of that internally reflected once only none is scattered through angles less than 30° unless $m < 1{\cdot}02$ or $m > 7{\cdot}66$. This is not necessarily true with non-spherical particles, but if they are not orientated the light so scattered will be mainly through wider angles. Internal reflection more than once will be relatively unimportant since the Fresnel reflection coefficients are low, and fall as m approaches unity unless the critical angle is approached or exceeded. Finally, scattering by irregular particles cannot become concentrated in particular directions as can that by spheres (as at the rainbows) or by assemblies of crystals in which relatively few interfacial angles preponderate. In the following discussion it will therefore be assumed that m is low, and that internal reflection can be neglected when considering the light scattered forward.

Stetter (1949) has pointed out that the projections on the surfaces of dust particles obscure very oblique surfaces, and so reduce external reflection at grazing incidence. However, if $m \simeq 1$ external reflection is less important than transmission and diffraction.

§ 4. APPARATUS

Only the light within the first dark ring of the Kirchhoff diffraction pattern is readily interpretable, and light scattering apparatus can be designed accordingly. Two types of measurement were made: (*a*) the light extinction was measured with different angles of acceptance of the transmitted light, (*b*) the laterally scattered light was measured directly ($7\tfrac{1}{2}° < \theta < 35°$ approximately). Only the former, however, is considered here.

The apparatus, shown diagrammatically in figure 1, was substantially the same as that employed by Vouk (1948), except that a brighter source was used (compact source mercury vapour lamp, filtered to give monochromatic green, blue or ultra-violet light). This permitted smaller apertures and better definition of θ_0, the scattering angle within which scattered light reaches the photocell.

In this method a parallel beam of light passes through one or other of two interchangeable, parallel-sided, glass cells. One of these contains the suspension, the other clear suspending liquid, and the light reaching the photocell is measured for each cell. The ratio of the two readings is the transmission T, and the angular distribution of light scattered forward is derived from the variation of T with distance between suspension and photocell, and hence with solid angle. There is no lens in the scattered beam, so θ_0 is ill-defined. It is influenced by (i) the divergence of the incident beam; (ii) the distance between suspension and photocell; and (iii) the ratio of the diameter of the illuminated volume to that of the aperture in front of the photocell. A lens in the scattered beam, with the aperture defining the scattered beam in its focal plane, eliminates all but the first of these (see, for example, Lothian and Chappel 1951), but apparatus of this type was not available.

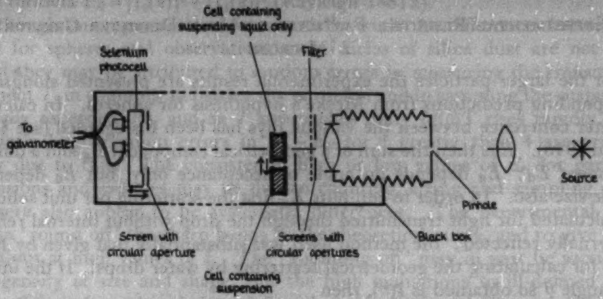

Cell containing suspending liquid only
Selenium photocell
Filter
To galvanometer
Source
Pinhole
Black box
Screen with circular aperture
Screens with circular apertures
Cell containing suspension

Figure 1. Diagram of apparatus for measuring extinction coefficient.

§ 5. Preparation of Suspensions

The dust suspensions employed were prepared by repeated sedimentation in liquid, using substantially the same technique as that used by Vouk (1948). The suspensions so prepared were far from monodisperse (see table), but were more homogeneous than the dusts obtainable from most sources.

Suspensions used in Extinction Measurements. Diameters in microns

Fraction	1	3	5	7	9	11	12
d_s	>14	7–10	3·5–5	1–2	0·4–0·6	0·2–0·3	<0·2
d_m	8–30	4–15	2·4–7·5	0·8–3·0	—	—	—
$(\overline{d_m^2})^{1/2}$	20·2	10·7	5·4	2·0	—	—	—

d_s = diameter of sphere with the same settling speed in the suspending medium used, d_m = diameter of sphere of the same projected area, measured microscopically; in each case the range quoted embraces about 95% of all particles sized. $(\overline{d_m^2})^{1/2}$ = diameter of sphere of same mean projected area.

The suspending liquid, both during separation and during the scattering measurements, was alcohol–water mixture (10% alcohol by volume, refractive index 4/3; for quartz in this medium $m = 1{\cdot}16$).

The area of cross section of the particles for the coarser fractions was measured by microscope, using a Patterson–Cawood graticule. Since particles settling on to a microscope slide under gravity tended to lie flat they were suspended in agar solution and 'frozen in' by rapid cooling (Ellison 1954). They were then sized as found in the gel. Numerical concentrations were determined by counting in cells of known depth. The finest fractions consisted of particles too small for accurate sizing by microscope.

§ 6. EXPERIMENTAL RESULTS: PARTICLES OF MEAN DIAMETER GREATER THAN 2 MICRONS

For the larger particles the experimental results are presented alongside the corresponding predictions from Mecke's hypothesis for spheres. In calculating the latter coherence between the various rays has been disregarded (see § 2, and Bricard 1946). E is then the sum of a geometrical component E_g and a diffracted component E_d; E_g depends on angle of acceptance only, but E_d depends on particle size also. In order to compute E_g first the scattering per unit solid angle was calculated for light transmitted through the drop without internal reflection or externally reflected: the method used was substantially that given by Bricard (1946) for calculating the geometrical scattering by water drops. If the intensity at an angle θ so obtained is $I(\theta)$, then

$$E_g = 1 - \int_{\theta=0}^{\theta_0} 2\pi I(\theta) \sin\theta \, d\theta,$$

and E_g was obtained by graphical integration. The contribution to E_g of light which is internally reflected was neglected, since for particles of as low relative refractive index as that of silica in water the Fresnel reflection coefficients are low. It is hoped to publish these computations subsequently.

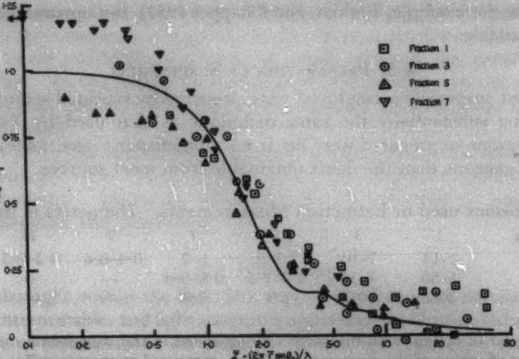


Figure 2. Difference between the measured extinction E_e and the extinction calculated according to geometric optics, E_g (mean particle size greater than 2 microns). The curve shows the corresponding difference for an opaque disc, calculated according to diffraction theory.

If the extinction found experimentally is denoted by E_e, figure 2 shows $E_e - E_g$ plotted as a function of $\bar{z}$, where $\bar{z} = 2\pi\bar{r}\sin\theta_0/\lambda$; $\bar{r}$ is the radius of the circle of the same mean projected area as that of the suspended particles as determined microscopically, and θ_0 is the conical half-angle suspended by the photocell at the centre of the suspension, measured in the suspending medium (diameter of aperture in front of photocell = 1·95 cm). The curve shows that part of the extinction arising from diffraction for monodisperse opaque discs, which by Mecke's hypothesis should be the same as E_d for spheres: it is given by the formula $E_d = [J_0(z)]^2 + [J_1(z)]^2$ (Rayleigh 1881 a).

Although this representation is essentially crude the differences between the theory for spheres and observations on particles of silica dust are not great. In part they may be attributed to random errors in measuring the transmission coefficient T, in counting and sizing the particles, and in sampling the suspension. T usually exceeded 0·6, and as T approaches unity $\delta E/\delta T$ rises rapidly. The apparently large random errors in $E_e - E_g$ for $\bar{z}$ greater than 4 are probably largely due to this. The consistently high or low values of E_e for particular suspensions and samples may be due to counting, sizing and sampling errors.

There also appear to be systematic differences. The experimental points seem to 'round off' the sharp bends in the theoretical curve, and to give higher extinctions at high values of $\bar{z}$. The 'rounding off' may in part be ascribed to heterogeneity of size and shape, and the high extinctions to ill-defined angles of cut-off θ_0 and rapid variation of E with θ_0 (Lothian and Chappel 1951).

For the finest suspension (fraction 7) $E_e - E_g$ consistently exceeds 1 for $\bar{z}$ less than 0·6, which corresponds to E_e exceeding 2. Both Mie's theory (see for example van de Hulst 1946) and, if coherence is taken into account, Mecke's approximation (Lothian and Chappel 1951), predict values of E fluctuating about a mean value rather greater than 2, as x increases, and rising well above 2 in the range $2 < 2x(m-1) < 6$. A very approximate extinction curve for spheres of refractive index 1·16 was therefore interpolated between Hart and Montroll's (1951) approximate curves for $m = 1{\cdot}25$ and $m \to 1{\cdot}00$. This was used to derive $\bar{E}$, a mean value of E weighted by area according to the formula $\bar{E} = \{\Sigma(nd^2E)\}/\{\Sigma(nd^2)\}$. So obtained $\bar{E} = 2{\cdot}46$ for green light and $\bar{E} = 2{\cdot}20$ for ultra-violet. These values correspond to zero acceptance angle, for which $\bar{z} = 0$ and $E_g = 1$, so that $\bar{E} - E_g$ approaches the values 1·46 and 1·20 respectively as $\bar{z}$ approaches zero. The latter agrees well with the experimental value shown in figure 2, and the former differs no more than might be expected. For the second finest suspension (fraction 5) a large part of the total cross-sectional area arises from particles of such sizes that $3{\cdot}2 < (m-1)x < 4{\cdot}6$ for green light or $6{\cdot}5 < (m-1)x < 7{\cdot}5$ for ultra-violet: these size ranges correspond to regions within which E falls below 2, and consequently $\bar{E}$ similarly computed for green light was not appreciably greater than 2 (Hart and Montroll's curves only extend to $(m-1)x = 8{\cdot}5$, so that $\bar{E}$ could not be calculated for ultra-violet). Moreover Hart and Montroll's approximation gives values of E which are too high at large values of x, so although the observed values of $E_e - E_g$ are less than 1 for this fraction, in this case the difference, considerably smaller than was found with fraction 7, is probably within the limits of experimental error.

§ 7. Experimental Results: Particles of Mean Diameter less than 1 Micron

For suspensions of these particles the particle size distribution cannot be determined accurately by optical microscope, so that E_e cannot be calculated. The angular distribution of the forward scattering was studied as before.

The sizes included in these suspensions lie near the boundary between Mecke and Rayleigh–Gans regions. Here the intensities and polarization predicted by the two approximations differ, but the angular distributions are very similar, and consequently discrimination between them by means of the transmission is possible only if either E or $y = -d(\log E)/d(\log \lambda)$ is known. For Rayleigh–Gans scattering $E \ll 1$ and $2 < y \leqslant 4$ (van de Hulst 1946: see also Atherton and Peters 1953), whereas for Mecke scattering the mean of E is about 2 and that of y is zero.

Between these two extremes the extinction of spheres shows marked peaks when plotted as a function of x, and y can assume either a positive or a negative value (LaMer 1948). The nature of the variation is known only for spheres of those refractive indices for which Mie's formulae have been computed (LaMer 1948). For non-spherical particles any estimate of size from $dE/d\lambda$ rests on assumptions regarding the shape and orientation of the particles.

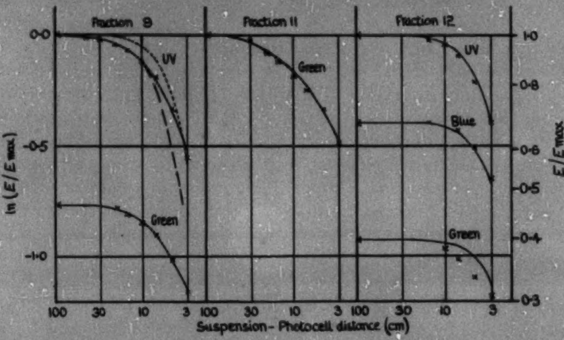


Figure 3. Dependence of extinction on distance between photocell and suspension (mean particle size less than 1 micron). The crosses indicate experimental results, and the curves indicate the values obtained by calculation from various postulated particle size distributions. Key:

Fraction 9	Rayleigh–Gans scattering,	=0·6 micron.	- - - - - - -
	Mecke scattering,	=0·88 micron.	— — — —
	2/5 Mecke scattering, 3/5 Rayleigh–Gans scattering,	=1·8 microns. =0·4 micron.	————
Fraction 11	1/2 Rayleigh–Gans scattering, 1/4 Mecke scattering, 1/4 Mecke scattering,	=0·4 micron. =2 microns. =6 microns.	————
Fraction 12	Rayleigh–Gans scattering,	=0·48 micron.	————

The insensitivity of forward scattering both to particle size distribution and to the choice of Mecke or Rayleigh–Gans hypothesis is shown in figure 3. The experimental results for the three finest fractions, and the theoretical results predicted on various assumptions, are shown. In order to be readily comparable the data are plotted on logarithmic scales. Since E is not known, the highest measured extinction for each suspension $E_{\max}$ is arbitrarily equated to 1 and all other measured extinctions are expressed in terms of it. The calculated extinction coefficients at zero angles of acceptance are equated to those observed at 175 cm for light of the same wavelength.

For fractions 9 and 11, and with suspension–photocell distances greater than 10 cm ($\theta_0 < 4°$), more light reaches the photocell than is predicted on the assumption of Rayleigh–Gans scattering by spheres of the same density and settling rate as the particles of the suspension. This suggests the presence of larger particles. (E, θ) curves calculated on the assumptions that the particle size distribution of fraction 9 has two peaks and that that of fraction 11 has three give much better agreement than those assuming monodisperse suspensions.

Microscopic examination of fractions 9 and 11 immediately after gentle agitation revealed no large single particles, but aggregates of approximately the sizes postulated above were present, and those in fraction 11 were larger than those in fraction 9. Vigorous shaking broke them up, but this was not possible immediately before scattering measurements because the bubbles formed clung to the cell walls. These aggregates were not sized since their size distribution was clearly unstable. With fraction 12, presumably because of lower numerical concentrations, very few aggregates were seen.

The presence of aggregates is not surprising, since high numerical concentrations are necessary if the scattering is to be measurable. Under the microscope, using dark field illumination, they appear compact and scatter diffusely, and the diffracted light may be assumed to behave in the same way as that diffracted by single particles of the same size. The number of aggregates, and the mean size of the individual particles, are clearly over-estimated by the coefficients used in deriving the 'theoretical' curves, since these relate to equivalent areas; for the aggregates $E \simeq 2$, whereas for the particles E is very much less than 1 and rises rapidly with x.

The presence of aggregates makes the slope y difficult to interpret. For the coarsest fraction $y = 1{\cdot}9$, which is below the lowest value for Rayleigh–Gans scattering. Since the Mie formula has not been computed no estimate of size is possible. For the finest fraction $y = 2{\cdot}3$. In the Rayleigh–Gans region this corresponds to $x = 7$ approximately. This seems large, but if $x = 7$, $2x(m-1) = 2{\cdot}24$, and the Rayleigh–Gans hypothesis does not strictly hold. Moreover, the result will be substantially weighted by the larger particles, for which $y \ll 2$.

§ 8. DISCUSSION

The flux scattered by the larger particles is approximately twice that incident on their geometrical cross section, but if $1 < d < 2{\cdot}5$ microns it exceeds this. These results agree with van de Hulst's predictions (1946) (see figure 2) that for transparent spheres the highest value of E occurs at about $\rho = 2x(m-1) = 4$, and that for $1{\cdot}5 < \rho < 6$ the extinction exceeds 2. For the two wavelengths used in most of the experiments, the condition $\rho = 4$ corresponds to diameters of approximately 1·5 and 1 micron respectively, so the agreement is good. Vouk (1948) found similar high values of E for coal particles, rising to 2·6 for particles of diameter 2 microns, and to 2·9 for particles of diameter 1·05 microns. This is, however, less easy to fit into the general picture drawn by van de Hulst (1949). Vouk quotes the relative refractive index of coal in water as being $m = 1{\cdot}4(1{\cdot}0 - 0{\cdot}3i)$. Putting this in the form used by van de Hulst, $m = n - ik = 1{\cdot}4 - 0{\cdot}42i$, and $\beta = \tan^{-1}(k/n-1) = 46°$ approximately, and for the 1·05 and 2·0 micron particles $\rho = 2(n-1)x$ has the values of 4·5 and 8·5 respectively. Two points arise: (i) although $\rho = 4{\cdot}5$ lies near the first peak in E for transparent spheres, van de Hulst's curves show that as the absorption k (and hence β) increases all peaks and troughs in the $E-\rho$ curve are damped out; if $\beta=45°$ E does not rise above 2; (ii) $\rho=8{\cdot}5$ corresponds to a range in which $E<2$ even for transparent particles. Admittedly van de Hulst's data for absorbing particles refer to the case where $n\rightarrow1$, and are scarcely applicable here, but the discrepancy seems large.

The experimental data relating to the smaller particles are less complete and less reliable, and their interpretation is complicated by the presence of aggregates, but the forward scattering also seems to resemble that by spheres. This is to be expected, since at 0° the intensity of Rayleigh–Gans scattering depends on the volume of the particle only (van de Hulst 1946), and only at wide angles does shape begin to affect it (Doty and Steiner 1950).

The experiments described above were made using suspensions of compact particles in which no one shape predominated, and only with such suspensions is the optical behaviour likely to resemble that of suspensions of spheres (see § 3).

§ 9. CONCLUSIONS

The experimental results show that, within the accuracy of the measurements made, the forward scattering of light by polydisperse suspensions of silica dust in liquid can be approximated to that by spheres of the same relative refractive index and the same particle size distribution. For the purposes of comparison, the latter has been calculated from the Mecke approximation for spheres of diameter greater than about 1 micron, and from the Rayleigh–Gans approximation for those smaller. For particles whose cross-sectional area is equal to that of spheres of diameter about 2 microns, the extinction is considerably greater than 2, in agreement with Mie's theory; for larger particles it is about 2, and for particles below 1 micron it was not measured. Owing to the formation of aggregates the scattering by particles smaller than 0·8 micron is through smaller angles than expected.

These conclusions seem likely to hold for any suspensions sufficiently polydisperse to conceal the maxima and minima of the Kirchhoff diffraction pattern, provided that no one shape predominates and the shapes are in the main not extreme.

These general features arise from the fact that forward-scattered light mainly corresponds to low-order Kirchhoff diffraction or to transmission through the particles, without internal reflection, in accordance with the laws of geometrical optics. At wider scattering angles and with higher relative refractive indices the diffraction will be of higher order and the internally reflected light may not be negligible. In this case the simplification may therefore distort the true picture.

ACKNOWLEDGMENTS

The author is very grateful to Dr. T. Bedford and Dr. C. N. Davies for advice and encouragement in the course of this work.

REFERENCES

ATHERTON, E., and PETERS, R. H., 1953, *Brit. J. Appl. Phys.*, **4,** 344.
BRICARD, J., 1946, *Ann. Géophys.*, **2,** 231.
BRILLOUIN, L., 1949, *J. Appl. Phys.*, **20,** 1110.
DOTY, P., and STEINER, R. F., 1950, *J. Chem. Phys.*, **18,** 1211.
ELLISON, J. MCK., 1954, *Nature, Lond.*, **173,** 948.
GANS, R., 1925, *Ann. Phys., Lpz.*, **76,** 29.

Hart, R. W., and Montroll, E. W., 1951, *J. Appl. Phys.*, **22**, 376.
LaMer, V. K., 1948, *J. Phys. Chem.*, **52**, 65.
Lothian, G. F., and Chappel, F. P., 1951, *J. Appl. Chem., Lond.*, **1**, 475.
Mecke, R., 1920, *Ann. Phys., Lpz.*, **62**, 623.
Mie, G., 1908, *Ann. Phys., Lpz.*, **25**, 377.
Montroll, E. W., and Greenberg, J. M., 1952, *Phys. Rev.*, **86**, 889.
Montroll, E. W., and Hart, R. W., 1951, *J. Appl. Phys.*, **22**, 1278.
Rayleigh, Lord, 1871, *Phil. Mag.*, 4th series, **41**, 447; 1881 a, *Ibid.*, 5th series, **11**, 214; 1881 b, *Ibid.*, 5th series, **12**, 81; 1910, *Proc. Roy. Soc.* A, **84**, 25: 1918, *Phil. Mag.*, 6th series, **35**, 373.
Sinclair, D., and LaMer, V. K., 1949, *Chem. Rev.*, **44**, 245.
Stetter, G., 1949, *Microtecnic* (English Edition), **3**, 234.
van de Hulst, C. H., 1946, *Rech. Astr. Obs. Utrecht*, XI, pt. 1; 1949, *Ibid.*, pt. 2.
Vouk, V., 1948, *Ph.D. Thesis*, University of London.

The Influence of Temperature on Metallic Wear

By J. K. LANCASTER

Research Laboratory, Associated Electrical Industries Limited, Aldermaston Court, Aldermaston, Berkshire

Communicated by W. Hirst; MS. received 10*th September* 1956

Abstract. The influence of temperature on the wear rate of a particular combination of metals, 60/40 brass on tool steel, has been investigated over the range 20°c to 600°c. Two regimes of wear were found. At low temperatures, extensive intermetallic contact and welding occur, and the wear rate increases with increasing temperature. At high temperatures, however, the wear rate decreases by several orders of magnitude and protective surface films are generated during sliding. The latter reduce both the extent of intermetallic contact and the frequency with which it occurs. A transition between these two regimes of wear also occurs at a critical load, the magnitude of which increases with the temperature. At loads above the transition load, the increase in wear rate with increasing temperature is accounted for by the associated decrease in hardness of the 60/40 brass, and the relationship $W=KL/P_m$ is obeyed, where W is the wear rate, L is the load and P_m is the flow pressure. At loads below the transition load, however, there is no definite relationship between the wear rate, load and flow pressure. It is suggested that in this regime of wear, the main effect of temperature arises from its influence on the reaction rate between the sliding metals and the surrounding atmosphere.

§ 1. INTRODUCTION

DIRECT proportionality between the wear rate of metals and the load applied during sliding has been predicted theoretically (Holm 1946, Burwell and Strang 1952, Archard 1953) but is not generally observed in practice (Hirst and Lancaster 1956, Archard and Hirst 1956). Archard and Hirst (1956) have recently suggested that the reason for this is that the surface conditions attained during sliding do not usually remain constant as the load is varied. A major factor influencing the surface conditions is the rise in temperature associated with frictional heating. An increase in temperature may change the chemical reactivity of the metals and affect the formation of protective surface films, and hence the degree of intermetallic contact. On the other hand, if it is assumed that wear is dependent in some way on the mechanical properties of the sliding metals, then a change in these properties with temperature could also affect the wear rate. The present paper is concerned with the way in which the temperature dependence of the above two factors influences the wear rate of a particular metal combination.

Previous investigations on the influence of temperature on wear have been limited to experiments in which an increase in temperature is produced as a result of frictional heating. In this way, Hughes and Spurr (1955) have studied the wear of Montan wax on cast iron. They find that although the wear rate is not proportional to the load, by measuring the surface temperature developed during sliding and determining the relevant value of the hardness of the wax, the relationship $W = KL/P_m$ is accurately obeyed where W is the wear rate, L is the applied load and P_m is the flow pressure, identified with the indentation hardness. This expression is the same as that predicted theoretically. Similarly, Sata (1953) has shown that although the wear rate of cast iron is not proportional to the load in experiments at constant speed, it becomes so when the speed is adjusted to maintain the same surface temperature at each load.

An alternative approach to the investigation of the effect of temperature on wear is to use a low speed of sliding in order to minimize frictional heating, and to increase the temperature by external heating. The experiments to be described have been made in this way. The metals used were 60/40 brass containing 2% lead sliding on hardened tool steel. Previous experiments have shown that with this combination the wear rate of the 60/40 brass is accurately proportional to the load over a wide range of loads (Archard 1953). The wear mechanism for this combination has also been investigated in detail (Kerridge and Lancaster 1956).

§ 2. EXPERIMENTAL

The apparatus used consisted of a 'pin and ring' type of wear machine and is shown schematically in figure 1. The base plate on which the loading arm and bearings were supported could be removed, to allow the insertion of the tube furnace, and then accurately replaced. Loads were applied by means of weights hanging vertically beneath the pin through a small hole in the base of the furnace. The rotating rings were made from Stag Special high speed tool steel, hardened to approximately 800 Vickers diamond pyramid number and ground to an external diameter of 1 in. The 60/40 brass pins, conforming to British Standards Specification 249, were in the form of ¼-inch diameter cylinders with a 45° semi-angle conical end.

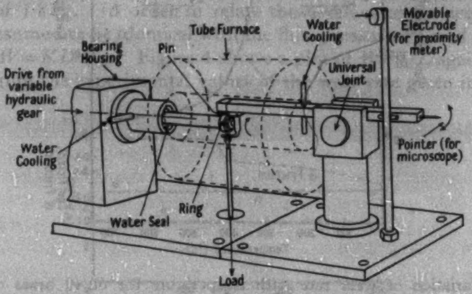


Figure 1. Schematic diagram of the wear machine.

The volume of metal worn from the pin was determined by measuring the decrease in depth of its conical end from microscope observations of a pointer attached to the loading arm outside the furnace. The loading arm was mounted in a universal bearing assembly, and the friction was measured from the deflections of a restraining beam detected with a capacitance proximity meter. The loading arm was also electrically insulated from the base plate in order to permit contact resistance measurements between the pin and the ring. Contact resistances were determined by measuring the potential drop across the contacts using a 1000 c/s a.c. supply and an applied e.m.f. of about 100 mv. Fluctuations were smoothed electrically, and the resistances obtained are therefore mean values over several repeated cycles of sliding. As already mentioned, the sliding speed was made as low as possible in order to minimize frictional heating; the actual value used was 1·3 cm sec^{-1}.

§ 3. Results

3.1. *Variation of Wear Rate with Temperature*

The changes in the wear rate of 60/40 brass occurring as the temperature increases from 20°c to 600°c are shown in figure 2. These results were obtained for a constant load of 2 kg, and the individual values of the wear rate were determined from the limiting slopes of the wear–time curves. There are two main regimes of wear. Below 300°c the wear rate increases with increasing temperature and the wear–time curves are either linear from the origin or slightly convex to the time axis as shown inset in figure 2. The wear phenomena in this region are characteristic of 'severe' wear involving the formation of a transferred film of brass on the tool steel ring and extensive intermetallic welding (Kerridge and Lancaster 1956). Above 300°c, the wear rate decreases abruptly and the type of wear–time curve changes to that usually associated with the formation of a protective surface film (Hirst and Lancaster 1956). This abrupt change in wear rate is also associated with the onset of an increase in contact resistance between the sliding metals. It is interesting to note, however, that the most rapid rate of increase in contact resistance occurs within the temperature range 400–450°c; within this range there is also a sharp increase in the coefficient of friction from 0·6 to 1·0.

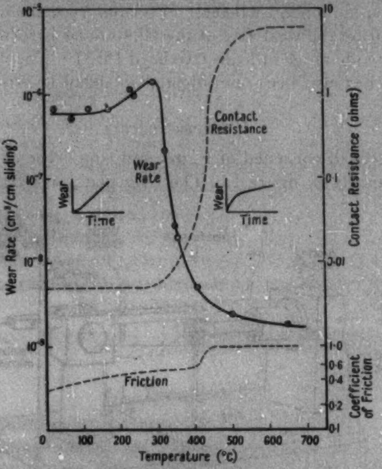


Figure 2. Variation of wear rate with temperature for 60/40 brass on tool steel.

3.2. *Variation of Wear Rate with Hardness*

The possibility that the increase in wear rate with temperature below 300°c is related to a decrease in hardness of the 60/40 brass was investigated by measuring the variation in hardness with temperature. Figure 3 shows the relationship obtained by Dr. B. Marriott of this laboratory. The hardness measurements were made using a standard Vickers diamond indenter in an argon atmosphere and at a load of 1·8 kg. In order to relate these values to the wear rates, the hardness was assumed to be identical with the flow pressure P_m and inserted into the expression $W = KL/P_m$. Figure 4 shows a plot of WP_m/L against temperature obtained from the experimental values of the wear rate given in figure 2.

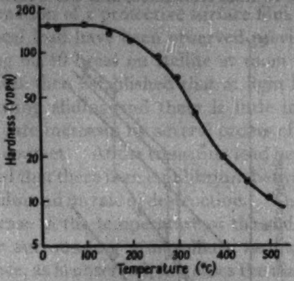


Figure 3. Variation of hardness with temperature for 60/40 brass.

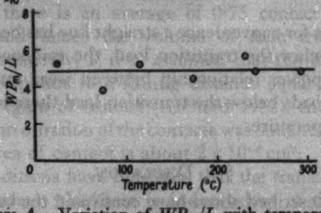


Figure 4. Variation of WP_m/L with temperature.

The above results show that the whole of the increase in wear rate below 300°c is accounted for by the variation in hardness with temperature, and that between 20°c and 120°c the hardness remains constant and the wear rate is independent of temperature. In experiments at 20°c, therefore, there will be a range of loads and speeds within which frictional heating is insufficient to produce surface temperatures in excess of 120°c. It is probably for this reason that the wear rate of 60/40 brass on tool steel has been found to be proportional to the load over a wide range of loads (Archard 1953).

3.3. *Variation of Wear Rate with Load*

The effect of temperature on the wear rate–load relationship is shown in figure 5. Two regimes of wear are obtained, similar to those already described in connection with the results given in figure 2, and the transition between these two regimes occurs at a critical load. The magnitude of this transition load increases with temperature from less than 200 g at 20°c to greater than 6 kg at 350°c. Above the transition load there is extensive intermetallic contact, and the wear rate is proportional to the load; below the transition load, however, there is a relatively high contact resistance, and the wear rate is not proportional to the load. Although for convenience a straight line has been drawn through the experimental points below the transition load, the evidence does not show that there is any definite power relationship between wear rate and load. It may also be noted that at loads below the transition load there is no regular variation in wear rate with temperature.

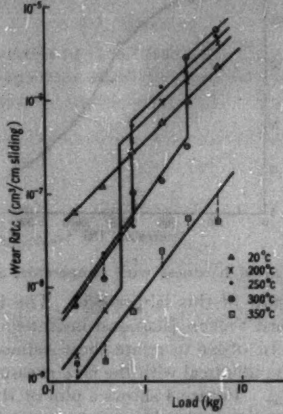


Figure 5. The effect of temperature on the wear rate–load relationship for 60/40 brass on tool steel.

§ 4. Discussion

The experiments described above have confirmed the validity of the relationship $W = KL/P_m$ for a severe type of metallic wear involving extensive intermetallic contact and welding. At constant temperature, and hence constant P_m, the wear rate is proportional to the load; at constant load the wear rate is inversely proportional to P_m. It follows, therefore, that with 60/40 brass on tool steel, the term K in the above expression is unaffected by changes in either the load or the flow pressure. Previous experiments have established that the 'severe' wear of 60/40 brass on tool steel occurs via the transfer of fragments of brass to the opposing tool steel surface (Kerridge and Lancaster 1956). Small localized intermetallic junctions grow and coalesce during sliding until the whole of the applied load becomes supported on a single region of contact, and the masses of the fragments transferred are proportional to the load divided by the hardness, i.e. to the real area of contact. If this process remains unaffected by temperature, and there is no obvious reason why this should not be so, the explanation for the increase in wear rate with temperature would be that larger fragments of metal are transferred rather than a greater number. The constancy of the term K therefore implies that the factors responsible for the initiation of intermetallic junctions and the production of transferred metal fragments are independent of the load and the flow pressure.

The abrupt decrease in wear rate obtained below a critical load is clearly associated with the formation of a protective surface film. Similar reductions in wear rate below a critical load have been observed previously during lubricated wear experiments using 60/40 brass on stellite at room temperature (Hirst and Lancaster 1956). It was then established that at light loads protective surface films are generated during sliding and there is little intermetallic contact; at heavier loads the wear rate increases by several orders of magnitude and there is extensive intermetallic contact. At the transition load between these two regimes of wear, it was suggested that there is an equilibrium between the rate of formation of a protective surface film and its rate of destruction. The present results support these views. An increase in the temperature of the sliding metals increases the reaction rate with the surrounding atmosphere, facilitates the formation of a protective film and hence, as is observed, increases the magnitude of the transition load.

Although the contact resistance measurements have shown that the degree of intermetallic contact is greatly reduced at loads below the transition load, this does not necessarily imply that intermetallic contact is of only minor significance. When a cathode-ray oscilloscope was connected across the sliding metals, it was found that intermittent decreases in resistance occurred from values of about 10 ohms to 0·5 ohm or less. Assuming that at each such decrease in resistance intermetallic contact occurs and a fragment of brass is transferred to the tool steel, the volume of the individual fragments may be calculated. Thus, at 650°c and a load of 2 kg, there is an average of 0·75 contact cm^{-1}, the wear rate is $1{\cdot}8 \times 10^{-9}\,\mathrm{cm}^3\,\mathrm{cm}^{-1}$ and hence the volume of each transferred fragment is $1{\cdot}6 \times 10^{-9}\,\mathrm{cm}^3$. Alternatively, if it is assumed that the duration of each intermetallic contact corresponds to a sliding distance equal to twice the diameter of a single circular region of contact (Archard 1953), the contact area may be calculated. The mean duration of the contacts was obtained from the oscilloscope, and the calculated area of contact is about $2 \times 10^{-6}\,\mathrm{cm}^2$. Previous experiments in 'severe' wear conditions have indicated that the transferred fragments are in the form of plates whose thickness is independent of the load (Kerridge and Lancaster 1956). Assuming that in the present conditions fragments of comparable thickness are produced, $2{\cdot}3 \times 10^{-4}\,\mathrm{cm}$, the volume of each fragment becomes $4{\cdot}6 \times 10^{-10}\,\mathrm{cm}^3$, of the same order of magnitude as the earlier estimate above.

The above calculations suggest that although intermetallic junctions occur only infrequently at loads below the transition load, they may, nevertheless, occur sufficiently often to be a possible main cause of wear. In the presence of a protective film, however, the area of intermetallic contact is much smaller than that which would be expected at the same load and temperature in the absence of a film ($2 \times 10^{-6}\,\mathrm{cm}^2$ compared with $2 \times 10^{-3}\,\mathrm{cm}^2$). The intermetallic junctions, therefore, do not grow to a size sufficient to support the load on a single contact region. The experimental results have shown that in this regime of wear the wear rate is neither proportional to the load nor inversely proportional to the flow pressure. In contrast to severe wear, therefore, in which the main effect of temperature is to change the hardness of the sliding metals, it is probable that the main effect of temperature below the transition load is upon the reaction rate between the metals and the surrounding atmosphere.

ACKNOWLEDGMENTS

The author wishes to express his thanks to Dr. W. Hirst for his continued interest and advice during the course of this work, to Mr. I. J. Webb for assistance with the experiments and to Dr. T. E. Allibone for permission to publish this paper.

REFERENCES

ARCHARD, J. F., 1953, *J. Appl. Phys.*, **24**, 981.
ARCHARD, J. F., and HIRST, W., 1956, *Proc. Roy. Soc.* A, **236**, 397.
BURWELL, J. T., and STRANG, C. D., 1952, *J. Appl. Phys.*, **23**, 18.
HIRST, W., and LANCASTER, J. K., 1956, *J. Appl. Phys.*, **27**, 1057.
HOLM, R., 1946, *Electrical Contacts* (Stockholm : Gerbers).
HUGHES, G., and SPURR, R. T., 1955, *Proc. Phys. Soc.* B, **68**, 106.
KERRIDGE, M., and LANCASTER, J. K., 1956, *Proc. Roy. Soc.* A, **236**, 250.
SATA, T., 1953, *Proc. 3rd Japan National Congress for Applied Mechanics*, 153.

Theory of Nuclear Magnetic Resonance in Eu^{3+}

By R. J. ELLIOTT
Physics Department, Reading University

MS. received 19*th September* 1956, *and in final form* 5*th November* 1956

Abstract. A calculation is made of the splitting of the energy levels of Eu^{3+} ions in their ground state which arises from the interactions of the nuclear magnetic moment with external and internal magnetic fields and of the nuclear quadrupole moment with the crystalline electric field. The feasibility of determining these nuclear properties by nuclear magnetic resonance is discussed. The method is applicable to all paramagnetic substances with singlet ground states.

---

§ 1. Introduction

Nuclear magnetic resonance is usually only observed in substances which are diamagnetic or only very weakly paramagnetic. In an ordinary paramagnetic ion, the electrons set up a magnetic field at the nucleus which may be of the order of 10^6 gauss and is much greater than the external field applied in a resonance experiment. Some substances, however, display at low temperatures a relatively weak paramagnetism which remains constant as $T \rightarrow 0$ (Van Vleck 1932). In such substances, the magnetic species have non-degenerate electronic ground states, which produce no magnetic field at the nucleus. It is the purpose of this paper to consider the possibility of observing nuclear resonance in such substances. The external magnetic field will induce a magnetic moment in the electrons which will in turn produce an internal field at the nucleus. As a result the position of the resonance will be shifted, an effect analogous to the 'chemical shift' often observed in molecules or to the Knight shift (Knight 1949) in metals. In view of the much larger paramagnetic susceptibility of this type of substance, an effect is to be expected larger by several orders of magnitude than these two effects. Higher order effects are also found to be important and will be considered.

Detailed calculations are presented for the Eu^{3+} ion, which is of interest because no direct measurement of nuclear moments by nuclear resonance has been possible in the rare earths, which do not form diamagnetic compounds. The nuclear spins have been determined from the hyperfine structure of paramagnetic resonance (Bleaney 1955, Bowers and Owen 1955) and less reliably from optical spectra. Nuclear magnetic moments can be determined from this hyperfine splitting if the electronic magnetic field responsible for it is calculated (Elliott and Stevens 1953, to be referred to as I): but the accuracy of such calculations depends on an estimate of the average value $\langle r^{-3} \rangle$, where r is the radius of the 4f electron. A direct measurement of one nuclear moment which, together with the hyperfine structure, would give $\langle r^{-3} \rangle$, would provide a basis for extrapolation to determine the other moments. Ytterbium does form diamagnetic compounds and is only weakly paramagnetic as a metal, so that it is a possibility. A determination of this information from europium may be difficult, for reasons outlined below.

These calculations were first performed some years ago and are now being published at the suggestion of Dr. B. Bleaney who recently began similar considerations into the feasibility of such an experiment on Eu^{3+} at Oxford. In other substances, similar calculations may be useful in the study of the electronic wave functions.

§ 2. EUROPIUM

2.1. *Magnetic Effects*

Eu^{3+} has a $4f^6$ configuration besides closed shells and, since Russell–Saunders coupling is an adequate approximation for our purposes, the lowest term is ^7F which is split by spin–orbit coupling into seven multiplets with $J=0, 1, \ldots, 6$. In the free ion the singlet $J=0$ is lowest with the triplet $J=1$ higher by energy Δ. The splitting of this level by a crystal field will vary from substance to substance and will be neglected for the moment. In the presence of an external field $\mathbf{H}$, the perturbing Hamiltonian is

$$\mathscr{H}=\beta\mathbf{H}\,.\,(\mathbf{L}+2\mathbf{S})+(2\beta\beta_N\mu_N/I)\langle r^{-3}\rangle\mathbf{N}\,.\,\mathbf{I}-(\mu_N\beta_N/I)\mathbf{H}\,.\,\mathbf{I}, \qquad \ldots\ldots(1)$$

where $\mathbf{L}$ and $\mathbf{S}$ are the total electronic orbital and spin angular momenta, $\mathbf{I}$ the nuclear angular momentum, β, β_N the electronic and nuclear Bohr magnetons, and μ_N the nuclear magnetic moment. $2\beta\mathbf{N}\,\langle r^{-3}\rangle$ is an operator defined in I (eqn 6.2) which gives the electronic magnetic field at the nucleus.

The splitting of the lowest level is obtained by perturbation theory using as basis the states $|J, M, m\rangle$, where M, m are the components of $\mathbf{J}$, $\mathbf{I}$ in the direction of the applied field. For H of the order of 10^3 gauss the three terms in equation (1) are in order of decreasing magnitude, but in first order only the third has an effect on this level giving the usual nuclear Zeeman energies

$$-(\mu_N\beta_N/I)Hm. \qquad \ldots\ldots(2)$$

Second-order perturbation theory gives an effect

$$-|\langle 0, 0, m|\mathscr{H}|1, 0, m\rangle|^2/\Delta. \qquad \ldots\ldots(3)$$

The largest contribution to this comes from the square of the first term in equation (1), but this is independent of m (as is the contribution of the square of the second term). The next largest comes from the cross product of the first two, namely:

$$-2\beta H\langle 0\|\Lambda\|1\rangle\ (2\beta\beta_N\mu_N/I)\langle r^{-3}\rangle\langle 1\|N\|0\rangle m/\Delta, \qquad \ldots\ldots(4)$$

where $\langle 0\|\Lambda\|1\rangle=2$ and $\langle 0\|N\|1\rangle=5/3$ are defined in I (p. 560, eqn (5.10), and p. 563). Equations (2) and (4) have the same form and the total splitting is therefore

$$-(\mu_N\beta_N/I)Hm(1-\alpha), \qquad \ldots\ldots(5)$$

where

$$\alpha=40\beta^2\langle r^{-3}\rangle/3\Delta=0{\cdot}89 \qquad \ldots\ldots(6)$$

for $\Delta=370\ \mathrm{cm}^{-1}$ (Sayre and Freed 1956) and $\langle r^{-3}\rangle=57\ \text{Å}^{-3}$ (Bleaney 1955). Thus the internal field set up by the electrons nearly cancels the external field. Effects in higher order perturbation theory are negligible.

The magnetic field of the electromagnetic radiation applied in the resonance experiment will also be screened by this factor α, so that the observed intensity will be reduced by $(1-\alpha)^2$. The smallness of this factor (6) introduces a serious difficulty into the experiment.

2.2. *Anisotropy*

In a crystal the ions are subjected to a crystalline electric field whose anisotropic nature produces a marked anisotropy in the resonance spectrum and would allow more interesting information to be obtained from an experiment. The effects are first examined for a field with an axis of symmetry z which is higher than two-fold. This field splits the $J=1$ level into a singlet $J_z=0$ and a doublet $J_z=\pm 1$ with energies Δ_0 and Δ_1, respectively, greater than $J=0$. Repeating the above calculation the Zeeman splitting is found to be

$$-(\mu_N\beta_N/I)[(\mathbf{H}\,.\,\mathbf{I})(1-\alpha_1)+H_zI_z(\alpha_1-\alpha_0)] \qquad \ldots\ldots(7)$$

where α_0, α_1 are defined by equation (6) with Δ_0, Δ_1 substituted. For the ethyl sulphate, Sayre and Freed (1956) give $\Delta_0=398\,\mathrm{cm}^{-1}$, $\Delta_1=356\,\mathrm{cm}^{-1}$, which makes the two terms in equation (7) roughly the same size. The intensity is still reduced, but by an anisotropic factor varying from $(1-\alpha_1)^2$ to $(1-\alpha_0)^2$ as $\mathbf{H}$ varies through 90° from the z-direction, assuming that the oscillating magnetic field is set as usual perpendicular to $\mathbf{H}$.

2.3. *Quadrupole Effects*

Because of the small magnetic energy, it is important to consider the effect of the nuclear electric quadrupole. The crystal field, already introduced, will interact with this directly, and also indirectly by distorting the electronic charge cloud of $J=0$ with admixtures of $J=2$ states. The relevant part of the crystal field has the form

$$V=A_2^0(3z^2-r^2) \qquad \ldots\ldots(8)$$

at (x, y, z), i.e. a constant A_2^0 times a second-order spherical harmonic. The direct interaction of this with the nuclear quadrupole Q (Casimir 1935) gives an energy

$$A_2^0Q[3I_z^2-I(I+1)]/(2I-1)I. \qquad \ldots\ldots(9)$$

The effect of the electronic distortion may be obtained by second-order perturbation theory and, if $|J=2, J_z=0\rangle$ has energy Δ_2, is

$$-2\langle 0, 0, I_z|A_2^0\sum_K(3z_K^2-r_K^2)|2, 0, I_z\rangle\langle 2, 0, I_z|\frac{e^2Q}{2I(2I-1)}\sum_K\frac{I(I+1)}{r_K^3}$$
$$-\frac{3\,\mathbf{r}(\kappa\,.\,I)^2}{r_K^5}|0,\ 0,\ I_z\rangle/\Delta_2, \qquad \ldots\ldots(10)$$

where Σ_K is over all f electrons. After some manipulation this reduces (see I, §§4, 6) to

$$\frac{2e^2Q\,A_2^0\langle r^2\rangle}{I(2I-1)\,\Delta_2}\langle r^{-3}\rangle|\langle 2\,||\alpha||\,0\rangle|^2[3I_z^2-I(I+1)]. \qquad \ldots\ldots(11)$$

$$\langle 2\,||\alpha||\,0\rangle=2/5\sqrt{3}\dagger.$$

A further 'pseudo-quadrupole' effect giving an energy proportional to I_z^2 arises from the magnetic terms—in fact from the square of the second term of equation (1) in expression (3)—and equals

$$5(\alpha_1-\alpha_0)\langle r^{-3}\rangle\beta_N^2\mu_N^2I_z^2/6I^2. \qquad \ldots\ldots(12)$$

In the particular case of the ethyl sulphate the separation of the two $J=1$ levels is

$$3A_2^0\langle r^2\rangle/5=42\,\mathrm{cm}^{-1}, \qquad \ldots\ldots(13)$$

giving $A_2^0\langle r^2\rangle=70\,\mathrm{cm}^{-1}$. Taking $\Delta_2=1015\,\mathrm{cm}^{-1}$ (Sayre and Freed 1956), the

---

† Private communication by Dr. B. Judd.

value (11) is $P(I_z^2-\frac{1}{3}I(I+1))$, where $P=1{\cdot}5Q\times 10^{-4}\,\mathrm{cm}^{-1}$ with Q in the usual units of $10^{-24}\,\mathrm{cm}^2$ and $I=5/2$ (Bleaney and Low 1955). The direct contribution (9) is about a hundred times smaller, and the pseudo-quadrupole effect (12) is of the order of $10^{-7}\,\mathrm{cm}^{-1}$. All effects in higher-order perturbation theory are negligible.

If one accepts the optical values of $Q=1{\cdot}2$ for isotope 151 and $Q=2{\cdot}5$ for 153 (Mack 1950), the level splittings are $4P$ and $2P$ where P is, respectively, 5 and 11 Mc/s, which gives transitions in a reasonable frequency region.

2.4. *Rhombic Symmetry*

Some crystals which might be more convenient for the experiment have crystalline fields of rhombic symmetry, e.g. $\mathrm{EuCl_3 . 6H_2O}$ (Hellwege and Kahle 1954). The $J=1$ level is now split into three states which transform like X, Y and Z, whose energies Δ_X, Δ_Y, Δ_Z may be determined by the use of polarized light. A straightforward extension of the above theory shows that the nuclear levels will be given by

$$-(\beta_N\mu_N/I)\sum_{i=X,Y,Z} I_iH_i(1-\alpha_i)+P(I_Z^2-\tfrac{1}{3}I(I+1))+R(I_X^2-I_Y^2), \quad ..(14)$$

where $\alpha_i=40\beta^2\langle r^{-3}\rangle/3\Delta_i$, $P=4e^2Q(2\Delta_Z-\Delta_X-\Delta_Y)\langle r^{-3}\rangle/15I(2I-1)\,\Delta_Z$

and $R=P(\Delta_X-\Delta_Y)/(2\Delta_Z-\Delta_X-\Delta_Y)$.

§ 3. CONCLUSION

If the anisotropic spectrum given by equations (7) and (11) or (14) were determined in a good single crystal the values of α_i and P would be found. Thus in a crystal where the splitting of the $J=1$ level, i.e. Δ_i, has been determined optically, μ_N, Q, and $\langle r^{-3}\rangle$ could all be obtained from the resonance data using the above theory. This is not possible without the anisotropic effects. A correction to the quadrupole moment Q should be applied to account for its polarization of the inner shells of electrons which has been shown to be several per cent for unionized europium (Sternheimer 1950, 1951). Polarization effects may also be important in determining the actual crystal field splitting (i.e. $A_2^0\langle r^2\rangle$) but the use of the empirical value will include these effects in calculating the quadrupole splittings (13) and (14).

The main difficulty in the experiment is the low intensity of the transitions, in which the nuclear moment appears to have less than 10% of its actual value. For the 151 isotope with the higher μ_N (by a factor of about 2·2) this apparent moment will probably be about 0·3, which is not impossibly small. In view of this difficulty it will be very important to have narrow lines. The width arising from the electronic magnetic moments on neighbouring ions will be greatly reduced at low temperatures, since the population of the $J=1$ level decreases like $\exp(-\Delta/kT)$ and the $J=0$ level has no moment. A neighbour in a $J=1$ state sets up a field of about 100 gauss, but the actual contribution is 1% of this at 80°K. The nuclear magnetic moments of surrounding nuclei set up a field which may be estimated from the line width of paramagnetic resonance on similar magnetically dilute crystals of other rare earths. The mean square is about 15 gauss in the ethyl sulphate, but is smaller in the double nitrates and anhydrous salts. Since all these fields are, however, also screened by the $1-\alpha$ factor, the lines should always be narrow—certainly below room temperature.

It may be concluded, therefore, that interesting information could be obtained from nuclear magnetic resonance on Eu^{3+}, though the experiment presents some difficulties.

ACKNOWLEDGMENT

I am much indebted to Dr. B. Bleaney for discussions of this problem, particularly concerning the feasibility of the experiment.

REFERENCES

BLEANEY, B., 1955, *Proc. Phys. Soc.* A, **68**, 937.
BLEANEY, B., and LOW, W., 1955, *Proc. Phys. Soc.* A, **68**, 55.
BOWERS, K. D., and OWEN, J., 1955, *Rep. Progr. Phys.*, **18**, 304 (London : Physical Society).
CASIMIR, H. B. G., 1935, *Physica*, **2**, 719.
ELLIOTT, R. J., and STEVENS, K. W. H., 1953, *Proc. Roy. Soc.* A, **218**, 553.
HELLWEGE, K. H., and KAHLE, H. G., 1954, *Z. Phys.*, **129**, 62.
KNIGHT, W. D., 1949, *Phys. Rev.*, **76**, 1260.
MACK, J. E., 1950, *Rev. Mod. Phys.*, **22**, 64.
SAYRE, E. V., and FREED, S., 1956, *J. Chem. Phys.*, **24**, 1216.
STERNHEIMER, R., 1950, *Phys. Rev.*, **80**, 102; 1951, *Ibid.*, **84**, 244.
VAN VLECK, J. H., 1932, *Theory of Electric and Magnetic Susceptibilities* (Oxford : University Press).

The Dependence of the Hall Coefficient of a Mixed Semiconductor upon Magnetic Induction as Exemplified by Indium Antimonide

By D. J. HOWARTH, R. H. JONES and E. H. PUTLEY
Radar Research Establishment, Great Malvern, Worcs.

MS. received 16*th August* 1956, *and in final form* 2*nd October* 1956

Abstract. The Bloch–Wilson theory for the Hall effect of a mixed conductor in an arbitrary magnetic field is discussed. A method of computation based upon the use of an electronic computer for fitting the theoretical expression to experimental data is described.

To test the theory the method was applied to InSb. Good agreement was obtained between the observed behaviour of two p-type single crystals and the theoretical expression. The fitting of the theory to the experimental results enables the concentration of holes and electrons and their mobilities to be determined over a range of temperature. This leads to values of the intrinsic carrier concentration which are in good agreement with those found by other workers. It is also shown that in these specimens the mobility ratio falls as the temperature is reduced and that this behaviour can be expected from the effects of lattice and impurity scattering on the holes and electrons.

---

§ 1. Introduction

CONSIDER a semiconductor containing free carriers in both the conduction band (electrons) and the full band (holes). Then (Chambers 1952) the Hall coefficient R will be

$$R=\frac{\sigma_1^2R_1+\sigma_2^2R_2+\sigma_1^2\sigma_2^2B^2R_1R_2(R_1+R_2)}{(\sigma_1+\sigma_2)^2+\sigma_1^2\sigma_2^2B^2(R_1+R_2)^2} \qquad \ldots\ldots(1)$$

where σ_1, R_1 are the conductivity in ohm cm^{-1} and Hall coefficient in cm^3 coulomb^{-1} associated with the carriers in the first band, σ_2, R_2 the corresponding quantities for the second band and B is the magnetic induction.

Wilson (1953, p. 236) has shown that the Hall coefficient of an extrinsic (single band) semiconductor obeying classical statistics and with the mean free path independent of energy is almost independent of magnetic induction. Thus,

$$\left.\begin{aligned} R&=\frac{3\pi}{8}\frac{1}{ne}, \qquad B\to 0\\ R&=\frac{1}{ne}, \qquad B\to\infty \end{aligned}\right\} \qquad \ldots\ldots(2)$$

a variation of about 18%. Equation (1) shows that even if R_1, R_2, σ_1, σ_2 were independent of B, there would be a variation of R with B. It will be shown later on that under certain conditions large variations of R can take place at inductions sufficiently low to make the changes in σ due to the magnetoresistance effect small.

Harman, Willardson and Beer (1954 a) have studied this effect in p-type InSb and p-type Ge. In the present paper an analysis similar to theirs has been used and applied to p-type InSb, but a more detailed correlation between theory and experiment has been obtained. This has been made possible by the use of an electronic calculator for making extensive numerical calculations, and by the use of a recording technique for measuring the Hall effect (Putley 1956) that has enabled results to be obtained in far greater detail than would be practicable with manually operated apparatus. It has been shown that in p-type specimens of a semiconductor which can be represented by the simple Wilson model of a mixed conductor, the concentration of holes and electrons and their mobilities can be calculated by adjusting the parameters in the theoretical expression for R to fit the experimental variation of R with B. The method is applicable over a range of temperatures near the point where the Hall coefficient changes sign, and for InSb this range is about 50°. By studying a number of specimens of different purity, it is thus possible to calculate the intrinsic carrier concentration over a wide range of temperature and, in particular, to determine it at lower temperatures than would be possible by direct measurement with the given specimen.

§ 2. THEORY

Referring again to equation (1), it will be noted that in an extrinsic specimen (either σ_1 or $\sigma_2=0$) the terms in B disappear, and in a pure intrinsic specimen ($R_1=-R_2$) they also disappear. In a p-type specimen, there will be a temperature at which $\sigma_1{}^2R_1=-\sigma_2{}^2R_2$, the usual condition for the Hall zero at weak induction, and it will be near this point that the dependence on B will be the greatest.

The electrical conductivity in the absence of a magnetic field is given by

$$\sigma=\sigma_1+\sigma_2=ne\mu_{\mathrm{n}}+pe\mu_{\mathrm{p}}. \qquad \ldots\ldots(3)$$

A straightforward application of the Bloch–Wilson theory of conduction applied to a two-band model (Wilson 1953, p. 234) gives the following formula for R, valid for an arbitrary transverse induction B (Harman, Willardson and Beer 1954 a, eqn. (13)):

$$R=\frac{3\sqrt{\pi}}{4pe}\,\frac{\{(nb^2/p)K_4(\gamma_{\mathrm{n}})-K_4(\gamma_{\mathrm{p}})\}}{\{(nb/p)K_5(\gamma_{\mathrm{n}})+K_5(\gamma_{\mathrm{p}})\}^2+\gamma_{\mathrm{p}}\{(nb^2/p)K_4(\gamma_{\mathrm{n}})-K_4(\gamma_{\mathrm{p}})\}^2} \qquad \ldots\ldots(4)$$

where

$$b=\frac{\mu_{\mathrm{n}}}{\mu_{\mathrm{p}}} \quad \ldots\ldots(5), \qquad \gamma_i=\frac{9\pi}{16}B^2\mu_i{}^2 \qquad \ldots\ldots(6)$$

$$K_\alpha(y)=\int_0^\infty\frac{x^\alpha\exp(-x^2)}{x^2+y}dx \qquad \ldots\ldots(7)$$

n and p are the concentrations of electrons and holes, and μ_{n} and μ_{p} are the respective mobilities.

The more important assumptions made in deriving (4) are three in number. First, the distribution function for both electrons and holes is the classical Maxwell–Boltzmann function. This appears to be a valid approximation as in most cases, in particular in InSb, degeneracy will only be important at highs temperatures (Harman, Willardson and Beer 1954 b), and should not be significant over the temperature range to be considered here. Secondly, the surfaces of constant energy in momentum space are spherical. Thirdly, the mean free paths of both electrons and holes are independent of energy. This will be true if thermal scattering predominates, but not if other sources of scattering, such as ionized impurity centres, are important. The last two assumptions will be discussed later in considering the application of this model to InSb.

Assuming that all the impurity levels are ionized, giving a concentration of extrinsic holes p_e independent of temperature, then

$$p_e = p - n. \qquad \ldots\ldots(8)$$

Equation (4) can then be written (cf. Harman *et al.* 1954 a, eqn. (17)):

$$R = -\frac{3\pi}{8p_e e}\,\frac{2\{(nb^2/p)K_4(\gamma_n) - K_4(\gamma_p)\}[1-(n/p)]}{\sqrt{\pi}\{(nb/p)K_5(\gamma_n) + K_5(\gamma_p)\}^2 + \gamma_p\{(nb^2/p)K_4(\gamma_n) - K_4(\gamma_p)\}^2}. \qquad \ldots\ldots(9)$$

For small inductions this reduces to the well known form

$$R_{B\to 0} = -\frac{3\pi}{8e}\,\frac{nb^2-p}{(nb+p)^2}. \qquad \ldots\ldots(10)$$

At temperatures in the exhaustion range, $n\to 0$, $p\to p_e$ and R is independent of temperature, and is given by

$$R_{B\to 0} = \frac{3\pi}{8p_e e}$$

(cf. equation (2)). Thus from measurements of R at low temperatures and low inductions, p_e can be found.

Equation (9) will be used to discuss the variation of R with B. Typical experimental results for InSb are shown by the points plotted in figures 9 and 10. It is seen that the variation is very large; at certain temperatures R changes sign with increasing field strength, and the magnitude of the variation can be several times the value at low fields.

Using equations (5) and (8) there are three unknowns in equation (9): n (or p), μ_n and μ_p (or b). Using the measured values of the conductivity at zero induction (equation (3)) and the Hall coefficient at low fields (equation (10)) two of these quantities can be eliminated. In this way a single parameter family of curves was obtained in terms of the parameter b. The curve which gave the best fit with the experimental data determined the value of b. The values of the carrier concentration and mobilities may then be obtained from equations (3), (5) and (10).

The computation of the expression (9) is complicated by the presence of the integrals $K_\alpha(y)$. These functions may be expressed in terms of error functions and exponential integrals and, for large y, asymptotic formulae (see Harman *et al.* 1954 a, eqns. (6)–(9)). In the present work, however, use has been made of an electronic computer, and the integrals $K_\alpha(y)$ have been determined by numerical integration of equation (7). The computation may be carried out at any temperature for which R varies significantly with B, and so, for a given sample, the carrier concentrations and mobilities may be calculated over a range of temperature. The calculation proves to be easily prepared for the computer, and therefore can be carried through at high speed.

Some typical results are shown by the curves drawn in figures [2], [3], [4], [9] and [10]. The curves are calculated from equation (9) with b chosen to give the best fit with the experimental data. The points plotted on the figures are the experimental ones. The agreement is good except at large inductions in the lower temperature range, where the theory predicts a numerically larger Hall coefficient than is observed. These results will be discussed in detail in a subsequent part of this paper on the application of this method to indium antimonide.

§ 3. APPLICATION TO INDIUM ANTIMONIDE

3.1. *Experimental*

Two p-type single crystals of InSb were measured over the temperature range 14° to 290°K. The variation of R and σ with temperature was measured using a recording technique similar to that described by Putley (1956). This enabled the variation of R with T to be measured quickly using a range of values of magnetic induction between 100 and 10 000 gauss. In addition to this, measurements were made in constant temperature baths which enabled the variation of R with B to be studied down to 25 gauss. The baths used were liquid nitrogen, ethylene, and methyl chloride and solid carbon dioxide in acetone. The variation of R with B at intermediate temperatures was obtained by taking sections across the families of (R, T) curves obtained from recordings made for different values of B.

The measurements were taken down to 14°K to study the scattering mechanisms in the specimen and to determine the impurity activation energies and the extrinsic carrier concentrations. The results obtained are given in table 2. The temperature range over which the variation of R with B was big enough to apply the analysis was 170–220°K for one specimen and 200–260°K for the other. In a subsidiary experiment the temperature of the Hall zero was measured. This of course is a function of induction, and over the range available the temperature could be changed by about 30° for both specimens, as shown in figure 5.

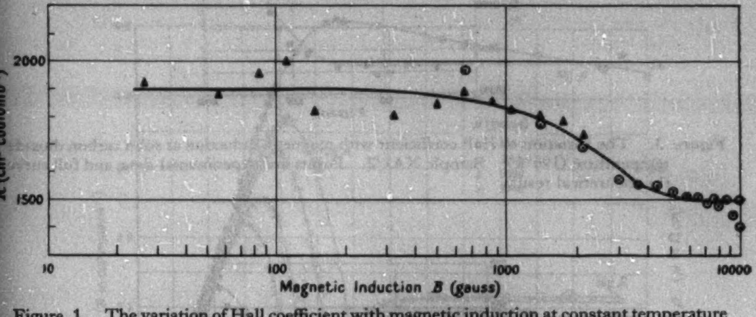


Figure 1. The variation of Hall coefficient with magnetic induction at constant temperature in the extrinsic range $T=77$°K. Sample XA1/2.

Figure 1 shows the variation of R with B at 77°K for specimen XA1/2. This is in the extrinsic range and the variation is of the order predicted by equation (2), although the behaviour goes over to the high field case at a lower induction than Wilson's theory predicts. Figures 2, 3 and 4 show results obtained from measurements in constant temperature baths. Figures 6 and 7 show the results obtained using the recorder. Log R is plotted against $1/T$ for a number of values of B, and log ρ the resistivity is also plotted. Figure 8 shows the behaviour of R and σ in the extrinsic range. Figures 9 and 10 give the points obtained by taking sections at constant temperature across the families of curves plotted in figures 6 and 7. In figures 2, 3, 4, 9 and 10 the points are the experimental values and the curves are those obtained from the analysis described in the last section.

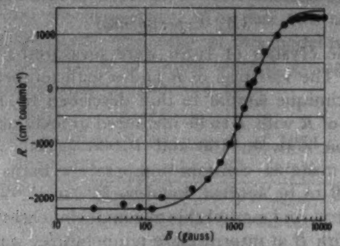


Figure 2. The variation of Hall coefficient with magnetic induction at liquid ethylene temperature (170°K). Sample XA1/2. Points are experimental data, and full curve the theoretical result.

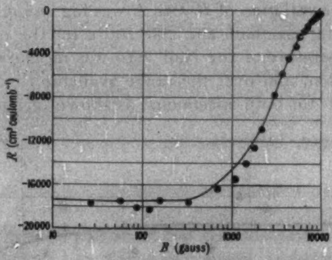


Figure 3. The variation of Hall coefficient with magnetic induction at solid carbon dioxide temperature (196°K). Sample XA1/2. Points are experimental data, and full curve the theoretical result.

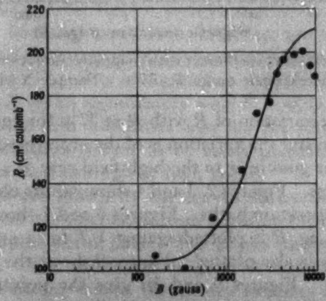


Figure 4. The variation of Hall coefficient with magnetic induction at solid carbon dioxide temperature (196°K). Sample XS1/1. Points are experimental data, and full curve the theoretical result.

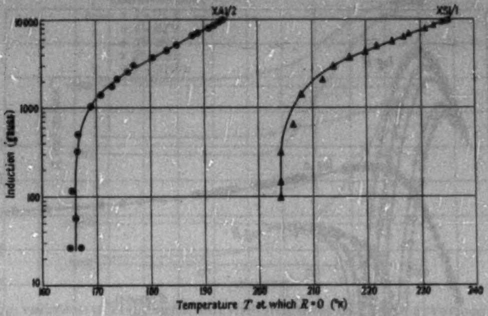


Figure 5. The effect of magnetic induction on the temperature of the Hall zero. Samples XA1/2 and XS1/1.

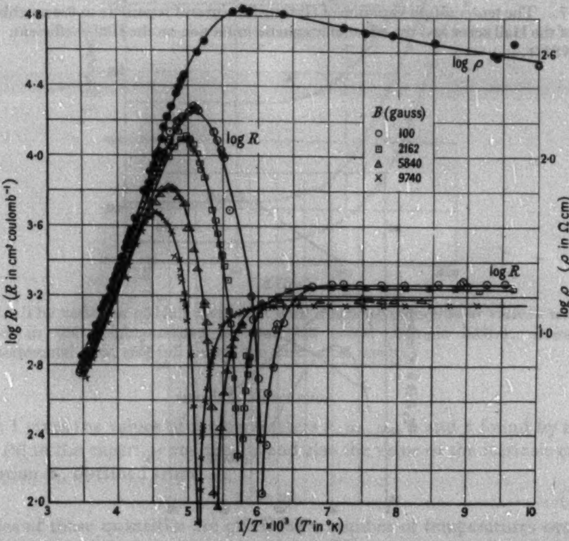


Figure 6. The temperature variation of Hall coefficient and resistivity in the neighbourhood of the Hall zeros and the effect of magnetic induction on the Hall coefficient. Sample XA1/2.

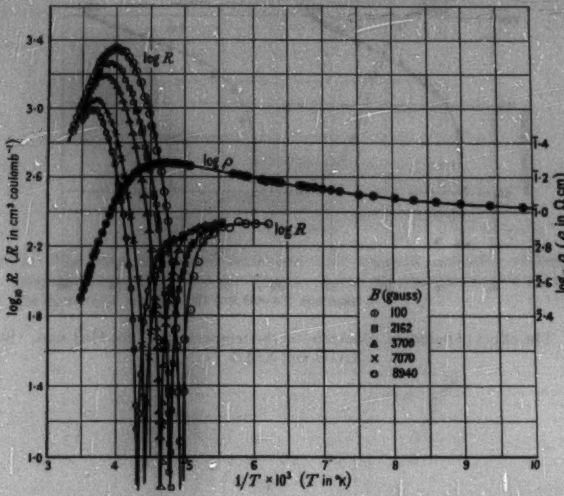


Figure 7. The temperature variation of Hall coefficient and resistivity in the neighbourhood of the Hall zeros and the effect of magnetic induction on the Hall coefficient. Sample XS1/1.

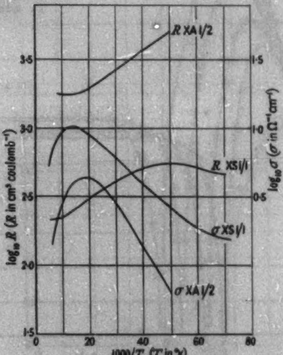


Figure 8. The extrinsic Hall coefficient and conductivity. Samples XA1/2 and XS1/1.

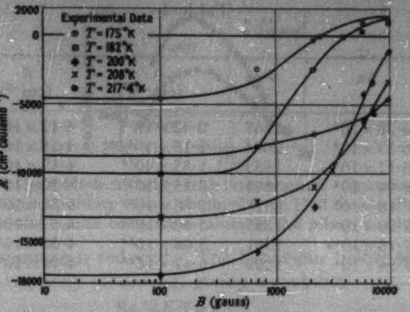


Figure 9. The variation of Hall coefficient with magnetic induction at various temperatures in the neighbourhood of the Hall zeros. Sample XA1/2. Points are experimental data, and full curves the theoretical results.

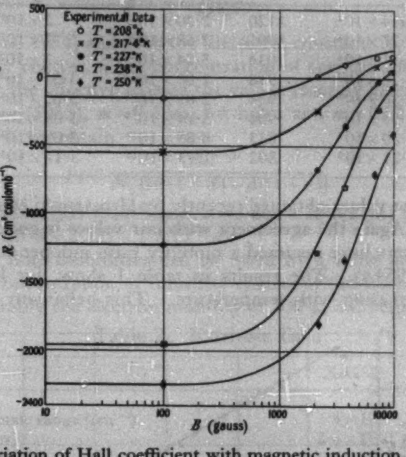


Figure 10. The variation of Hall coefficient with magnetic induction at various temperatures in the neighbourhood of the Hall zeros. Sample XS1/1. Points are experimental data, and full curves the theoretical results.

3.2. *Discussion*

Table 1 gives the values of the parameters b, μ_n, μ_p, n and p found by fitting equation (9) to the experimental results and also the value of the intrinsic carrier concentration n_i, obtained from

$$n_i^2 = np. \qquad \text{......(11)}$$

The values of these quantities are given for a number of temperatures over the range for which the calculation could be made. The ranges for the two specimens overlapped, and it will be seen that the values of n_i obtained for both specimens agree very well in the overlapping region. Figure 11 is a plot of log n_i against $1/T$

Table 1

Sample 1 : XA1/2

T	b	μ_n	μ_p	n (cm^{-3})	p (cm^{-3})	n_i (cm^{-3})
164	29	$6{\cdot}28\times10^4$	2167	$0{\cdot}42\times10^{13}$	$4{\cdot}17\times10^{15}$	$1{\cdot}61\times10^{14}$
169·5	30	$6{\cdot}13\times10^4$	2044	$1{\cdot}15\times10^{13}$	$4{\cdot}18\times10^{15}$	$2{\cdot}19\times10^{14}$
175	31·5	$6{\cdot}15\times10^4$	1953	$1{\cdot}85\times10^{13}$	$4{\cdot}19\times10^{15}$	$2{\cdot}78\times10^{14}$
182	36·3	$6{\cdot}86\times10^4$	1891	$3{\cdot}35\times10^{13}$	$4{\cdot}20\times10^{15}$	$3{\cdot}76\times10^{14}$
200	47	$7{\cdot}18\times10^4$	1528	$1{\cdot}88\times10^{14}$	$4{\cdot}36\times10^{15}$	$9{\cdot}04\times10^{14}$
208	52	$7{\cdot}20\times10^4$	1385	$3{\cdot}60\times10^{14}$	$4{\cdot}53\times10^{15}$	$1{\cdot}28\times10^{15}$
213	55	$7{\cdot}16\times10^4$	1301	$5{\cdot}08\times10^{14}$	$4{\cdot}68\times10^{15}$	$1{\cdot}54\times10^{15}$
217·4	58	$7{\cdot}54\times10^4$	1299	$6{\cdot}67\times10^{14}$	$4{\cdot}94\times10^{15}$	$1{\cdot}82\times10^{15}$

Sample 2 : XS1/1

T	b	μ_n	μ_p	n (cm^{-3})	p (cm^{-3})	n_i (cm^{-3})
196	28	$3{\cdot}14\times10^4$	1120	$2{\cdot}30\times10^{13}$	3×10^{16}	$8{\cdot}31\times10^{14}$
200	30	$3{\cdot}27\times10^4$	1091	$2{\cdot}68\times10^{13}$	3×10^{16}	$8{\cdot}97\times10^{14}$
208	31	$3{\cdot}21\times10^4$	1034	$5{\cdot}24\times10^{13}$	3×10^{16}	$1{\cdot}25\times10^{15}$
213	33	$3{\cdot}26\times10^4$	988	$1{\cdot}06\times10^{14}$	$3{\cdot}01\times10^{16}$	$1{\cdot}80\times10^{15}$
227	36	$3{\cdot}36\times10^4$	934	$2{\cdot}05\times10^{14}$	$3{\cdot}02\times10^{16}$	$2{\cdot}49\times10^{15}$
238	39	$3{\cdot}53\times10^4$	904	$3{\cdot}74\times10^{14}$	$3{\cdot}04\times10^{16}$	$3{\cdot}37\times10^{15}$
250	39·5	$3{\cdot}30\times10^4$	833	$8{\cdot}85\times10^{14}$	$3{\cdot}09\times10^{16}$	$5{\cdot}23\times10^{15}$
263	40	$3{\cdot}21\times10^4$	802	$1{\cdot}73\times10^{15}$	$3{\cdot}17\times10^{16}$	$7{\cdot}40\times10^{15}$

and on this plot the values obtained recently by Hrostowski *et al.* (1955) are given for comparison. Again the agreement with our values is good.

Previous workers have assumed a mobility ratio independent of temperature (Harman *et al.* 1954 a). The results in table 1 show, for both specimens, a mobility ratio increasing with temperature. This behaviour can be accounted for by considering the effect of ionized impurity scattering. Figure 12 shows the values of $R\sigma=\mu_p$ obtained in the extrinsic range, and also the values of μ_p obtained at slightly higher temperatures by the curve fitting. It is seen that above about 40°K μ_p falls as the temperature rises, but at lower temperatures

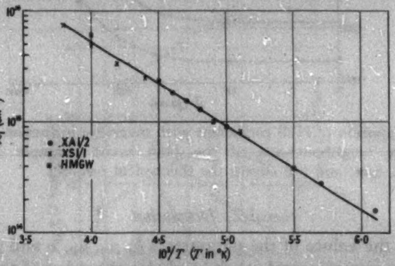


Figure 11. The variation of intrinsic electron concentration with temperature. Samples XA1/2 and XS1/1. Also included are the results of Hrostowski *et al.* (1955).

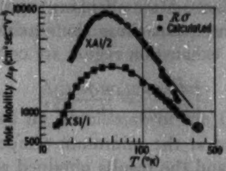


Figure 12. The variation of hole mobility with temperature. Samples XA1/2 and XS1/1. The squares denote mobility values obtained from Hall data in the extrinsic temperature range and the circles calculated values in the mixed conduction range.

it falls as the temperature is reduced. The behaviour below 40°K approximates to Brook–Herring impurity scattering, giving a mobility

$$\mu_{\mathrm{I}}=\frac{8\sqrt{2K^2(\boldsymbol{k}T)^{3/2}}}{\pi^{3/2}N_{\mathrm{I}}e^3m^{1/2}[\ln(1+d)-d/(1+d)]}\ (\text{e.s.u.}), \qquad \ldots\ldots(12)$$

where

$$d=\frac{6}{\pi}\frac{Km\boldsymbol{k}^2T^2}{nh^2e^2}$$

(Debye and Conwell 1954) and K is the dielectric constant, N_{I} the concentration of ionized impurity centres, n the concentration of carriers and m their effective mass. Taking for the effective masses the values recently given by Dresselhaus, Kipp and Wagoner (1955), $m=0{\cdot}18m_0$ for holes and $m=0{\cdot}013m_0$ for electrons, and putting $K=14$, we obtain

$$\mu_{\mathrm{I}}=\frac{AT^{3/2}}{N_{\mathrm{I}}[\ln(1+d)-d/(1+d)]}$$

where for holes $A=1{\cdot}517\times10^{18}$ and $d=3{\cdot}264\times10^{14}(T^2/n)$ and for electrons $A=5{\cdot}645\times10^{18}$ and $d=2{\cdot}357\times10^{13}(T^2/n)$.

The observed hole moblilities can be fitted by a value of N_{I} of about 10^{16} (see table 2). A concentration of impurity centres of this magnitude is sufficient to reduce considerably the electron mobility at 200°K, although not having much

Table 2. Extrinsic Data

	XA1/2	XS1/1
Excess holes in extrinsic range (cm^{-3})	4×10^{15}	3×10^{16}
Slope (log R, $1/T$) (ev)	$2{\cdot}6\times10^{-3}$	$2{\cdot}2\times10^{-3}$
Slope (log σ, $1/T$) (ev)	$5{\cdot}6\times10^{-3}$	$3{\cdot}6\times10^{-3}$
N_{I} estimated at 30°K	$9{\cdot}3\times10^{15}$	$5{\cdot}2\times10^{16}$
N_{I} estimated at 20°K	12×10^{15}	$6{\cdot}2\times10^{16}$
Mobility ratio calculated from $R_{\min}$	44	44
Temp. for $R_{\min}$ (°K)	197	250
n_{i} at temp. for $R_{\min}$	$6{\cdot}45\times10^{14}$	$4{\cdot}5\times10^{15}$

effect on the holes above 50°K. Thus in the region in which b has been determined by the application of equation (9) we should expect the mobilities of electrons and holes to behave very differently as the temperature falls. The hole mobility will increase while that of the electron will tend to fall. This behaviour will account for the variation of b with temperature found from equation (9).

It is also of interest to compare the values of b found from the application of equation (9) with those found by calculation using only the low induction values of the Hall coefficient. After changing sign the Hall coefficient passes through a minimum before approaching the intrinsic range. The minimum value, and the value in the exhaustion region are connected by the relation

$$\left|\frac{R_{\mathrm{ex}}}{R_{\mathrm{min}}}\right| = \frac{4b}{(b-1)^2}, \qquad \ldots\ldots(13)$$

where the value of b is that at the minimum temperature. This relation was applied to both specimens and the results obtained are given in table 2. Again it is seen that there is good agreement with the values given in table 1. It is also possible to calculate a value for n_i at the Hall minimum by using the relation

$$\frac{R_{\mathrm{min}}}{R_i} = \frac{b+1}{4\sqrt{b}}, \qquad \ldots\ldots(14)$$

where

$$R_i = -\frac{3\pi}{8n_i e}\frac{b-1}{b+1}$$

in conjunction with equation (13). The results are also given in table 2 and agree with the values found in table 1.

Equation (12) was used to calculate N_I from the observed values of $R\sigma$ at 20°K and 30°K. The value obtained at 20°K was about 20% higher than that found at 30°K. This suggests that some other mechanism was reducing the mobility. Neutral impurity scattering (Erginsoy 1950) could not account for this, but the presence of impurity band conduction is suspected because below 20°K R passes through a low maximum similar to those observed by Fritzsche and Lark–Horovitz (1955).

§ 4. CONCLUSION

It has been seen that the method outlined for dealing with the variation of the Hall coefficient with magnetic induction can be applied successfully to indium antimonide, although some discrepancies between theory and experiment have been observed. The values of R for large induction tended to be smaller than the predicted values. In some cases R went through a maximum at about 6000 gauss and then decreased as B was raised further (see figure 4).

The presence of impurity scattering may explain this deviation between theory and experiment. This leads to a mean free path depending on energy as E^2 and any deviation from a mean free path independent of energy should be more important at higher induction since R depends on B through products of B with the mobility. Since the scattering by impurity centres will increase in importance as the temperature is lowered, we should expect this discrepancy to become more important at lower temperatures. The theory outlined assumes the constant energy surfaces in space to be spherical. Available evidence (Herman 1955) suggests that this is a valid approximation for the conduction band in InSb, but more complicated structures have been suggested for the valence band (Hrostowski *et al.* 1955). The failure of this theory to account for the behaviour in Ge has been attributed to the complex nature of its valence band (Harman *et al.* 1954 a). Although the theory takes no direct account of band shape, Wilson (1953, § 8.22) has pointed out that a two-band theory can be expected to explain, at least qualitatively, results due to non-spherical energy surfaces since it involves energy surfaces with both positive and negative curvatures. For InSb the model appears to describe correctly the main features of the results.

For those semiconductors to which this method can be applied it provides a useful means for calculating the carrier concentrations and mobilities in the mixed conduction range.

ACKNOWLEDGMENT

This paper is published by permission of the Controller, H.M. Stationery Office.

REFERENCES

CHAMBERS, R. C., 1952, *Proc. Phys. Soc.* A, **65**, 903.
DEBYE, P., and CONWELL, E., 1954, *Phys. Rev.*, 93, 693.
DRESSELHAUS, G., KIPP, A. F., and WAGONER, G., 1955, *Phys. Rev.*, **98**, 556.
ERGINSOY, C., 1950, *Phys. Rev.*, 79, 1013.
FRITZSCHE, H. and LARK-HOROVITZ, K., 1955, *Phys. Rev.*, 99, 400.
HARMAN, T. C., WILLARDSON, R. K., and BEER, A. C., 1954 a, *Phys. Rev.*, 95, 699; 1954 b, *Ibid.*, 93, 912.
HERMAN, F., 1955, *J. Electronics*, 1, 103.
HROSTOWSKI, H. J., MORIN, F. J., GEBALLE, T. H., and WHEATLEY, G. H., 1955, *Phys. Rev.*, 100, 1672.
PUTLEY, E. H., 1956, *J. Sci. Instrum.*, 33, 164.
WILSON, A. H., 1953, *Theory of Metals* (Cambridge : University Press).

On the Propagation of High Frequency Compressional Waves in Isotropic Cylinders

By M. REDWOOD and J. LAMB

Electrical Engineering Department, Imperial College, London

MS. received 8*th October* 1956

Abstract. The possible modes of propagation of compressional waves in solid cylindrical rods are described by the solutions of the 'frequency equation'. It has been found possible to solve this equation at frequencies near 10 Mc/s by a graphical method which is readily applicable. Using this method the variation of phase velocity v_p with frequency has been evaluated for the most important modes of propagation, namely those having phase velocities near to that in an unbounded medium, v_l. It is shown that in the vicinity of $v_p = v_l$ the variation of phase velocity with frequency is characterized by a series of 'plateau' regions over which v_p is approximately constant, these regions being separated by 'steps' through which v_p changes rapidly with frequency.

Consideration of the displacement amplitudes shows that the mode with v_p nearest to v_l has, in general, a pressure variation across the radius which is most likely to be excited experimentally under conditions described in a previous paper by the authors.

---

§ 1. Introduction

It is possible at frequencies of about 1 Mc/s and above to propagate a compressional wave along the axis of a cylindrical rod in one or more different modes. At a particular frequency each mode is described by a certain phase velocity and pressure distribution over the cross section. The characteristics of these modes follow from a solution of the so-called 'frequency equation' which for a given rod relates the possible phase velocities to frequency. Although the solutions of the frequency equation at low frequencies are well known (Bancroft 1941, Davies 1948) only two sets of calculations have been published at high frequencies and both referred to a fixed frequency of 10 Mc/s (Adem 1954, McSkimin 1956).

The aim of the present paper is to present a method for graphical solution of the frequency equation in the high-frequency region. This enables a rapid evaluation to be made of the roots of the equation over a narrow frequency range centred on a selected frequency. This yields both the phase velocities and the pressure distributions over a cross section: these form the basis of a general understanding of propagation in cylindrical rods.

§ 2. Analysis of Mode Propagation

2.1. *Basic Equations*

The equations of motion of an isotropic elastic solid with no body forces present are expressed in the form:

$$\rho\partial^2\mathbf{u}/\partial t^2 = (\lambda+2\mu)\nabla\nabla\cdot\mathbf{u} - \mu\nabla\times\nabla\times\mathbf{u} \qquad \ldots\ldots(1)$$

where $\mathbf{u}$ is the vector displacement; λ and μ are Lamé's constants and ρ is the density. In cylindrical polar coordinates (r, θ, z) the three stress components acting radially may be expressed:

$$\begin{aligned} \sigma_{rr} &= \lambda\Delta + 2\mu\partial u_r/\partial r \\ \sigma_{r\theta} &= \mu\left\{\left[\frac{1}{r}\frac{\partial u_r}{\partial \theta}\right] + r\frac{\partial(u_\theta/r)}{\partial r}\right\} \\ \sigma_{rz} &= \mu\left\{\left[\frac{\partial u_r}{\partial z}\right] + \frac{\partial u_z}{\partial r}\right\} \end{aligned} \qquad \ldots\ldots(2)$$

where u_r, u_θ and u_z are the components of the vector displacement $\mathbf{u}$ and the dilatation $\Delta = \nabla \cdot \mathbf{u}$.

The boundary conditions require that the three normal stresses shall be zero at the surface of the cylinder $r = a$ so that each equation in (2) is identically equal to zero at $r = a$.

Let an infinite train of sinusoidal compressional waves of frequency $f = \omega/2\pi$ be propagated in the z direction parallel to the axis of the cylinder. Then $u_\theta \equiv 0$ and the remaining displacements may be written:

$$u_r = U_r \exp j(\gamma z - \omega t), \quad u_z = U_z \exp j(\gamma z - \omega t). \qquad \ldots\ldots(3)$$

Confining our attention to axially symmetrical waves, U_r and U_z are functions of r *only* and γ is the propagation constant: the phase velocity $v_{\mathrm{p}} = \omega/\gamma$.

It follows from the above that U_r and U_z must take the form (Love 1927)

$$\begin{aligned} U_r &= A\frac{\partial}{\partial r}\mathrm{J}_0(\alpha r) + C\gamma \mathrm{J}_1(\beta r) \\ U_z &= Aj\gamma \mathrm{J}_0(\alpha r) + \frac{Cj}{r}\frac{\partial}{\partial r}[r\mathrm{J}_1(\beta r)] \end{aligned} \qquad \ldots\ldots(4)$$

where $\alpha^2 = [\rho\omega^2/(\lambda+2\mu)] - \gamma^2 = h^2 - \gamma^2 = \omega^2(1/v_{\mathrm{l}}^2 - 1/v_{\mathrm{p}}^2)$

and $\beta^2 = [\rho\omega^2/\mu] - \gamma^2 = K^2 - \gamma^2 = \omega^2(1/v_{\mathrm{t}}^2 - 1/v_{\mathrm{p}}^2)$. ......(5)

Here v_{l} and v_{t} are the velocities of compressional (longitudinal) and transverse (shear) waves, respectively, in an unbounded medium.

Application of the boundary conditions now gives the 'frequency equation' (Love 1927)

$$\begin{vmatrix} 2\mu\dfrac{\partial^2}{\partial a^2}\mathrm{J}_0(\alpha a) - \dfrac{\lambda}{(\lambda+2\mu)}\omega^2\rho \mathrm{J}_0(\alpha a), & 2\mu\gamma\dfrac{\partial}{\partial a}[\mathrm{J}_1(\beta a)] \\ 2\gamma\dfrac{\partial}{\partial a}\mathrm{J}_0(\alpha a), & (2\gamma^2 - \omega^2\rho/\mu)\mathrm{J}_1(\beta a) \end{vmatrix} = 0 \qquad \ldots\ldots(6)$$

where $\partial/\partial a$ means $\partial/\partial r$ taken at $r \equiv a$.

2.2. *Consideration of the Frequency Equation* (6).

The most convenient form of the frequency equation is that used by Adem (1954); if the radius a is put equal to unity equation (6) can be rewritten

$$[\gamma^2\beta \mathrm{J}_0(\beta)/\mathrm{J}_1(\beta)] - \tfrac{1}{2}K^2 + (\tfrac{1}{2}K^2 - \gamma^2)^2 \mathrm{J}_0(\alpha)/\alpha \mathrm{J}_1(\alpha) = 0. \qquad \ldots\ldots(7)$$

The phase velocities of the possible modes of propagation at any particular frequency are given by the roots of this equation. The behaviour at low frequencies is well known (Davies 1948, Bancroft 1941, Holden 1951), the variation of phase velocity with frequency being given in figure 1 for the first three modes. At low frequencies mode 1 is the Young's modulus vibration having a velocity tending to the value $(E/\rho)^{1/2}$ as f tends to zero. At high frequencies the phase velocity of this mode approaches, asymptotically, the velocity of Rayleigh surface waves v_s (Davies 1948).

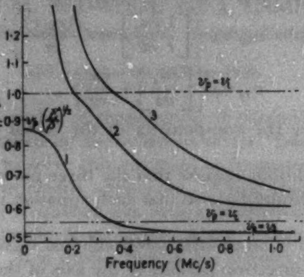


Figure 1. Phase velocity (v_p) of compressional waves in a cylindrical bar at low frequencies (*after* Davies 1948).
Poisson's ratio $\nu=0{\cdot}29$; radius of rod=1 cm; $v_l=5\times10^5$ cm sec^{-1}.

Present interest centres on the propagation of compressional waves in a cylindrical rod at frequencies of the order of 10 Mc/s. Thus, we are concerned with the higher order modes which exhibit a cut-off frequency (such as modes 2 and 3 of figure 1), and particularly with those modes which have a phase velocity close to v_l at the frequency considered.

Some attention has been given previously to this problem, notably by Holden (1951), Adem (1954) and McSkimin (1956) and the behaviour has been described qualitatively by Junger (1955). Numerical computations of the roots of the frequency equation were made by McSkimin and Adem for a frequency of 10 Mc/s but no investigation has previously been made of the dependence of phase velocity on frequency in this high-frequency region. Numerical solution of the roots of the frequency equation by successive approximations is laborious but, with certain assumptions, it is possible to use a graphical method which will now be described.

The frequency equation can be rewritten (after Adem 1954)

$$[h^2y^2\beta J_0(\beta)J_1(\beta)]-\tfrac{1}{2}K^2+(\tfrac{1}{2}K^2-h^2y^2)^2J_0(\alpha)/\alpha J_1(\alpha)=0 \qquad \ldots\ldots(8)$$

where $y=\gamma/h$. It is convenient to consider a material with a Poisson's ratio ν of 1/3 for then $v_l=2v_t$ and $K=2h$. Under these conditions equation (8) becomes

$$[y^2\beta J_0(\beta)/J_1(\beta)]-2+h^2(2-y^2)^2J_0(\alpha)/\alpha J_1(\alpha)=0. \qquad \ldots\ldots(9)$$

Also $$\alpha=h(1-y^2)^{1/2},\quad \beta=h(4-y^2)^{1/2}. \qquad \ldots\ldots(10)$$

Equation (9) is an equation in two variables, y and h: y is related to the phase velocity by $v_p=v_l/y$ and h is related to the frequency by $h=\omega/v_l$. Thus at any particular frequency h is fixed and the equation may be solved for y.

It is readily possible to determine the number of possible modes at any one frequency from the form of the frequency equation since the cut-off frequencies are determined from the condition $y=0(v_p=\infty)$, that is by the roots of the equations

$$hJ_0(h)/J_1(h)=\tfrac{1}{2} \quad\text{and}\quad J_1(2h)=0. \qquad \ldots\ldots(11)$$

It is also possible to determine how many of these modes have phase velocities *below* the 'free space' value v_l since the condition $v_p = v_l\ (y = 1)$ occurs once for each mode. The corresponding frequencies are given by the roots of the equation

$$J_1(h\sqrt{3}) = 0. \qquad \ldots\ldots(12)$$

2.3. *Graphical Determination of the Roots of the Frequency Equation*

Referring to the frequency equation (9), it was found from preliminary numerical calculations made for a frequency of 10 Mc/s that for the roots of interest (v_p near v_l) y is very nearly unity. With v_l equal to 5×10^5 cm sec^{-1} typical values of y at a frequency of 10 Mc/s for the first three modes which have phase velocities above v_l are $y = 0{\cdot}9998$, $0{\cdot}9989$ and $0{\cdot}9975$. These give values of β large compared with unity (> 200 cm^{-1}) and values of α between 0 and 10 cm^{-1}. It was realized that in order to gain a full appreciation of the behaviour detailed attention would have to be given to a restricted frequency region. This arises because the separation between the frequencies for which $v_p = v_l$ for successive modes is of the order of 150 kc/s, as determined by the roots of equation (12). Therefore, confining our attention to the frequency range of $(10 \pm 0{\cdot}15)$ Mc/s we found from the initial computations that it was possible to make certain approximations, namely:

(i) $y^2 = 1$

(ii) $\beta = 217$ cm^{-1} for $v_l = 5 \times 10^5$ cm sec^{-1}

(iii) $h^2 = 15\,800$ cm^{-2}.

The frequency equation (9) then becomes

$$[J_0(\beta)/J_1(\beta)] - 0{\cdot}0092 + 73J_0(\alpha)/\alpha J_1(\alpha) = 0. \qquad \ldots\ldots(13)$$

(iv) The middle term in this equation can safely be ignored, leading to the reduced frequency equation

$$J_0(\beta)/J_1(\beta) = -73J_0(\alpha)/\alpha J_1(\alpha). \qquad \ldots\ldots(14)$$

It must be noted that when $y > 1$ ($v_p < v_l$), α is imaginary and the equation becomes

$$\frac{J_0(\beta)}{J_1(\beta)} = +\frac{73\,I_0(|\alpha|)}{|\alpha|\,I_1(|\alpha|)}. \qquad \ldots\ldots(15)$$

In the graphical solution of these two forms of the frequency equation (14) and (15) it is convenient to make the further approximations

$$\text{(v)}\quad \beta = h\sqrt{3} + \alpha^2/2h\sqrt{3} = h\sqrt{3} + \alpha^2/435, \qquad \ldots\ldots(16)$$

which follows from $\beta^2 = 3h^2 + \alpha^2$.

$$\text{(vi)}\quad v_p = v_l(1 + \tfrac{1}{2}\alpha^2/h^2) = v_l(1 + \tfrac{1}{2}\alpha^2/15\,800), \qquad \ldots\ldots(17)$$

which follows from $v_p^2 = v_l^2/y^2 = v_l^2(1 - \alpha^2/h^2)^{-1}$.

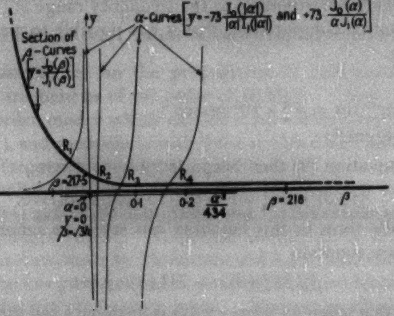
The approximate frequency equations (14) and (15) are solved in the following manner.

I. The function $J_0(\beta)/J_1(\beta)$ is plotted on an expanded β-scale against β for the range $\beta = 215$ to 225 cm^{-1}. (The actual length of the abscissa scale is approximately 15 feet). This is designated the β-curve.

II. To the same scale the functions $-73J_0(\alpha)/\alpha J_1(\alpha)$ and $+73I_0(|\alpha|)/|\alpha|I_1(|\alpha|)$ are plotted against $\alpha^2/435$ on transparent paper: these are designated the α-curves.

III. The α-curves are set arbitrarily over the β-curve to give a common abscissa. This fixes the frequency value for which the equations are to be solved for y, since $\beta = h\sqrt{3} + \alpha^2/435$ and $f = hv_1/2\pi$. The frequency in question is given by the value on the β-abscissa corresponding to the origin of the α-curves. Then the α and β values associated with this frequency are read off from the intersections of the two sets of curves. The value of the phase velocity v_p at a particular intersection may then be calculated from the corresponding value of α determined above using equation (17). Values of v_p derived in this way give the phase velocities at this frequency of those modes which have v_p close to v_l (assumption (i)).

In practice a direct scale of velocity is placed on the abscissa of the α-plot and a direct scale of frequency on that of the β-plot: the frequency at which a particular velocity value occurs is read off after suitable positioning of the α-curve. This procedure is illustrated in figure 2 and is then repeated until a sufficient number of points is obtained over the selected range of frequency. The results of this analysis are given in figure 3 which shows the general pattern of behaviour.



y
α-Curves $\left[y = -73\frac{I_0(|\alpha|)}{|\alpha| I_1(|\alpha|)} \text{ and } +73\frac{J_0(\alpha)}{\alpha J_1(\alpha)}\right]$
Section of β-Curves $\left[y = \frac{J_0(\beta)}{J_1(\beta)}\right]$
R₁
R₂
R₃
R₄
$\beta = 217.5$
$\alpha = 0$
$y = 0$
$\beta = \sqrt{3}h$
0·1
0·2
$\frac{\alpha^2}{434}$
$\beta = 218$
β

Figure 2. Graphical method of solution of the frequency equation at high frequencies. Thick lines refer to the β-curves (a section only of these curves is shown). Thin lines refer to the α-curves. The roots of the equation are given by the values of α at the intersections R_1, R_2, R_3, R_4. Note: y here has no relationship to the y of equation (8).

The dispersion curve for each mode of propagation shows several regions or plateaux over which the phase velocity is sensibly constant with changing frequency: these 'plateau' regions are joined by 'steps' through which the phase velocity changes markedly with frequency over a relatively small range of frequency. This behaviour applies to any mode until, with increasing frequency, its phase velocity has fallen in a succession of such steps to a value equal to the velocity v_l of free-space longitudinal waves. Beyond this point the phase velocity exhibits a pronounced decrease with increasing frequency and eventually approaches the value of the shear-wave velocity in an unbounded medium v_t. It can be seen from figure 3 that the curve for mode B is nearly flat for frequencies from 10·01 Mc/s to 10·13 Mc/s and that for mode A from 9·87 to 9·99 Mc/s. A similar behaviour is to be found at higher frequencies. Thus over a considerable portion of the frequency range there is a possible mode of propagation for which the phase velocity is sensibly constant at a value slightly greater than the free-space velocity v_1 (by about 2 parts in 10^4 for the case considered above).

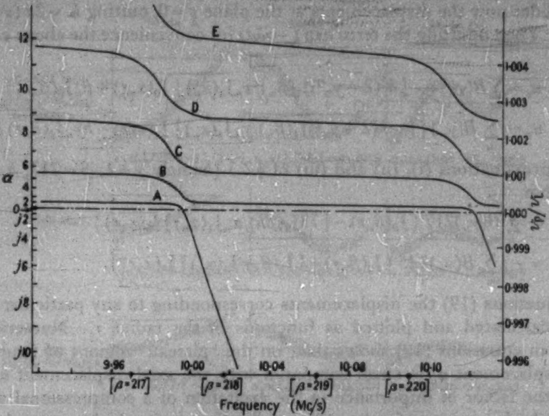


Figure 3. Phase velocity of compressional waves in a cylindrical bar at high frequencies. $\nu=1/3$, radius of rod $=1$ cm, $v_1=5\times10^5$ cm sec^{-1}. The 'plateaux' correspond to values of α close to the roots of $J_0(\alpha)=0$. The 'steps' correspond to values of β close to the roots of $J_1(\beta)=0$.

The regions of near-constant phase velocity are of importance in velocity measurements, especially those utilizing a pulse technique. Negligible dispersion of pulses will occur if conditions are such that it is possible to generate predominantly that single mode the phase velocity of which is nearest the free-space velocity v_1. This implies that none of the major frequency components of the pulse shall fall in the region of the 'steps' which occur at the extremities of the plateau. It should be noted, however, that even if these conditions are fulfilled the measured velocity is not v_1 but a value slightly higher than v_1. Dispersion will occur if more than one mode is generated owing to the differences in the phase velocities of the various modes: the velocities of successive modes may differ by 5 parts in 10^4 or more. Consideration will later be given to some practical aspects relating to the propagations of predominantly single modes.

2.4. *The Displacement Functions*

The radial and axial components of the displacement for any particular mode are given by equations (3) and (4). If the simultaneous excitation at frequency $f(=\omega/2\pi)$ of all possible modes is considered for some arbitrary driving force at $z=0$, the resulting displacements may be written

$$u_z=\sum_n B(\gamma_n)\left\{-\left[\frac{(\frac{1}{2}K^2-\gamma_n^{\,2})}{\alpha_n}\times J_1(\beta_n)/J_1(\alpha_n)\right]J_0(\alpha_n r)+\beta_n J_0(\beta_n r)\right\}\exp j(\gamma_n z-\omega t)$$

$$u_r=j\sum_n B(\gamma_n)\left\{-\left[\frac{(\frac{1}{2}K^2-\gamma_n^{\,2})}{\gamma_n}\times J_1(\beta_n)/J_1(\alpha_n)\right]J_1(\alpha_n r)-\gamma_n J_1(\beta_n r)\right\}\exp j(\gamma_n z-\omega t) \quad \ldots\ldots(18)$$

where $B(\gamma_n)$ is an amplitude function which is prescribed by the displacements applied at the plane $z=0$, Σ_n represents the total number of possible modes at the frequency in question, and the radius of the cylinder is taken as unity.

Consider now the displacements at the plane $z=0$, putting $K=2h$ $(\nu=1/3)$ as before. Then omitting the term $\exp(-j\omega t)$ for convenience the above equations become

$$u_z=\sum_n B(\gamma_n)\{-[h^2(2-y_n^2)\mathrm{J}_1(\beta_n)/\alpha_n\mathrm{J}_1(\alpha_n)]\,\mathrm{J}_0(\alpha_n r)+\beta_n\mathrm{J}_0(\beta_n r)\}$$

$$u_r=j\sum_n B(\gamma_n)\{[-h(2-y_n^2)\mathrm{J}_1(\beta_n)/y_n\mathrm{J}_1(\alpha_n)]\,\mathrm{J}_1(\alpha_n r)-hy_n\mathrm{J}_1(\beta_n r)\}.$$

With approximations (i), (ii) and (iii) of §2.3 (namely $y=1$, $\beta=217$, $h=125$) it follows that

$$\left.\begin{aligned} u_z&=\sum_n B(\gamma_n)217\{\mathrm{J}_0(\beta_n r)-[73\mathrm{J}_1(\beta n)/\alpha_n\mathrm{J}_1(\alpha_n)]\,\mathrm{J}_0(\alpha_n r)\}\\ u_r&=-j\sum_n B(\gamma_n)125\{\mathrm{J}_1(\beta_n r)+[\mathrm{J}_1(\beta_n)/\mathrm{J}_1(\alpha_n)]\,\mathrm{J}_1(\alpha_n r)\}.\end{aligned}\right\} \quad \ldots\ldots(19)$$

Using equations (19) the displacements corresponding to any particular mode n may be calculated and plotted as functions of the radius r. Numerical substitution in equations (19) shows that, on the 'plateau' regions of figure 3, the radial displacement u_r is small compared with the axial displacement u_z. The latter is the factor of importance in the excitation of a compressional wave as, for example, by means of a transducer vibrating in its thickness mode. The radial variation of u_z for modes A, B and C of figure 3 is illustrated in figure 4 for selected frequencies. The form of u_z across a radius is mainly determined by the term in $\mathrm{J}_0(\alpha_n r)$ in equation (19), with a superimposed 'ripple' due to the $\mathrm{J}_0(\beta_n r)$ term. The $\mathrm{J}_0(\alpha_n r)$ term is dominant for frequencies on the plateau regions but in the region of the steps the term in $\mathrm{J}_0(\beta_n r)$ becomes more important. The behaviour shown in figure 4 for mode A on its plateau is typical for any other mode when the frequency is such that its phase velocity lies on that plateau which is nearest to v_1.

§ 3. DISCUSSION

The results of the analysis given in §2 have important practical applications in the measurement of the velocity of propagation and the attenuation of compressional waves in solids. The usual ultrasonic transducers vibrating in a thickness mode produce, ideally, a uniform pressure distribution over the cross section at the plane of application. This pressure distribution does not coincide with that associated with any single mode of propagation so that, in practice, several modes are excited leading to dispersion due to mode interference. However, it has been shown in a previous paper (Redwood and Lamb 1956) that 'wringing' the transducer to the specimen with oil as a coupling medium produces a film which is thicker in the centre than at the periphery. This results in a pressure distribution over the cross section of the specimen very similar to that illustrated in figure 4A, and therefore to the propagation of a predominantly single mode. This explains why in the experimental work it was possible to eliminate dispersion due to mode interference and to make accurate determinations of the attenuation. It is also now evident that at the frequency of operation this single mode is the one which has a plateau in its dispersion curve nearest to the free-space velocity v_1. It follows that the measured velocity is somewhat greater than v_1, by about 2 parts in 10^4 at a frequency of 10 Mc/s. (It may be noted that this velocity difference is reduced by working at higher frequencies, or at the same frequency with rods of larger diameter.)

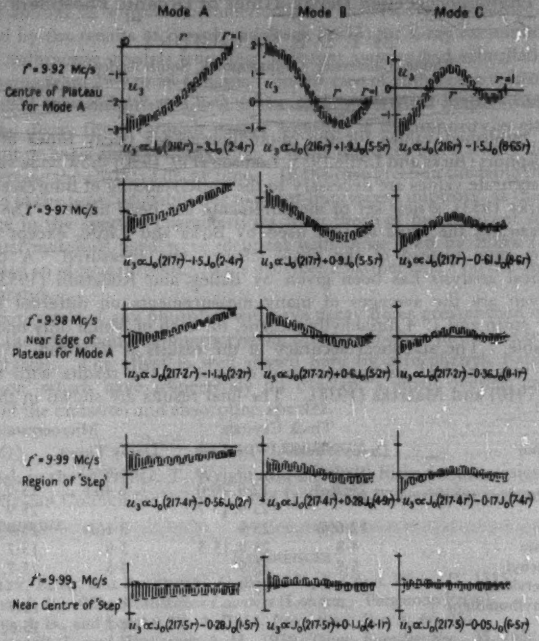


Figure 4. Variation of axial displacement u_z with radius at frequencies near 10 Mc/s. The $J_0(\beta r)$ term ($\beta \simeq 217$) produces a ' ripple ' on the displacement function: the ripple amplitude is nearly constant except near $r=0$, and goes through approximately 150 maxima and minima between $r=0$ and $r=1$. The limits of this ' ripple ' are shown by the broken lines: the ripple itself is represented by vertical shading.

ACKNOWLEDGMENTS

The authors are grateful for the financial support received from the Worshipful Company of Clothworkers for a research programme of which this work forms a part. M. R. wishes to acknowledge the financial assistance provided by the Oliver Lodge Scholarship of the Institution of Electrical Engineers.

REFERENCES

ADEM, J., 1954, *Quart. Appl. Math.*, **12**, 261.
BANCROFT, D., 1941, *Phys. Rev.*, **59**, 588.
DAVIES, R. M., 1948, *Phil. Trans. Roy. Soc.* A, **240**, 375.
HOLDEN, A. N., 1951, *Bell Syst. Tech. J.*, **30**, 956.
JUNGER, M. C., 1955, *J. Acoust. Soc. Amer.*, **27**, 974.
LOVE, A. E. H., 1927, *The Mathematical Theory of Elasticity*, (Cambridge: University Press).
MCSKIMIN, H. J., 1956, *J. Acoust. Soc. Amer.*, **28**, 484.
REDWOOD, M., and LAMB, J., 1956, *Proc. Instn Elect Engrs*, Pt. B, **103**, 773.

RESEARCH NOTES

Photofluorescence Decay Times of Organic Phosphors

By T. D. S. HAMILTON †
Rhodes University, Grahamstown, South Africa

MS. received 1*st November* 1956, *and in final form* 9*th November* 1956

The published values for the photofluorescence decay times of organic phosphors (Birks and Little 1953, Liebson *et al.* 1950) show wide variations. As accurate values are necessary to the understanding of fluorescence processes (Birks 1953), a new set of measurements has been made. The method used is basically the same as that used by Birks and Little, except that the modulation ratio instead of the phase change was measured. A complete mathematical analysis has been given by Bailey and Rollefson (1953). The results given are the averages of many measurements on different samples, except for eosin and 1: 4 diphenylbutadiene, for each of which only one sample was available. The statistical accuracy of the results is to within 2%. The apparatus was checked using eosin and comparing the results with those of Kirchoff (1940) and Maercks (1938). The final results are shown in the table.

	Thick Crystals		Microcrystals	
Phosphor	Decay Times (mμsec)	f (Mc/s)	Decay Times (mμsec)	f (Mc/s)
Anthracene	24·2(*a*)	12·9, 30·0 25·9, 15·7	6·4(*c*)	15·7, 30·0 25·9
Anthracene	12·6(*b*)	25·9	3·1(*d*)	25·9
trans-Stilbene	4·8	25·9, 15·5	3·9	15·7
para-Terphenyl	5·5	25·9	5·5	15·7, 30·0
Diphenylacetylene	4·9	25·9	—	—
1 : 4 Diphenylbutadiene	—	—	5·4	15·7
Eosin (*e*)	—	—	4·7	15·5, 25·9 30·0

(*f*) is the light modulation frequency; (*a*) newly cleaved surface (b) old surface long exposed to the atmosphere; (*c*) crystallized from spectroscopically pure solvent; (*d*) crystallized from contaminated solvent as used by Birks and Little; (*e*) solution in water.

The term ' thick crystal ' refers to 1 cm cube crystals, and ' microcrystals ' to approximately 0·1μ layers deposited from solution. The materials were supplied by the Larco Nuclear Instrument Company, Palisades Park, New Jersey, except for eosin from Distillation Products Industries–Eastman Kodak Company.

These results should be compared with those in the above references. Considering anthracene, it seems evident that the variations can be attributed to contamination of the surface of the crystal due to exposure to the atmosphere with probable oxidation (Bowen and Williams 1939). These results show how important it is to use clean surfaces for all photofluorescence measurements and scintillation measurements involving excitation with a very short range in the phosphor. This surface-layer effect also provides evidence in favour of a short-range energy-transfer process as discussed by Wright (1955). The decay may,

---

† Now at Department of Natural Philosophy, University of Glasgow.

however, not be exponential, in which case different results could possibily be obtained at different modulation frequencies. In view of the range of frequencies used here with identical results, this possibility may be excluded. The other possible cause of some of the differences is the presence of impurities even in the 'pure' materials as supplied commercially for scintillation counting. This is illustrated by the results of Swank and Buck (1955) for x-ray excitation. They find that anthracene crystals from four different sources had scintillation decay times ranging from 29·9 to 38·4 mμsec. In the case of anthracene microscrystals, impurities in the solvent remain with the anthracene after evaporation and can reduce the decay time (through energy transfer) if the impurity is either non-fluorescent or has a shorter decay time. If the impurity has a longer decay time it will increase the effective decay time. In the case given in the table there was a reduction of the time, probably due to a non-fluorescent impurity, as there was no difference in the spectral characteristics from those of pure solvent specimens. The actual materials used were identical with those used by Birks and Little, and yielded an identical result. This provides a further check on the operation of the apparatus.

Thick crystals of any phosphor will have decay times greater than or equal to those of the corresponding microcrystals, due to self-absorption and re-emission in the crystal. The effect is well shown here by *p*-terphenyl, *t*-stilbene and anthracene, which have respectively no overlap, a small overlap and a large overlap of the emission and absorption spectra.

ACKNOWLEDGMENTS

I wish to thank Dr. G. T. Wright and Dr. J. B. Birks for discussions, and the South African Council for Scientific and Industrial Research for maintenance and research grants.

REFERENCES

BAILEY, E. A., and ROLLEFSON, G. K., 1953, *J. Chem. Phys.*, **21**, 1315.
BIRKS, J. B., 1953, *Scintillation Counters* (London : Pergamon Press).
BIRKS, J. B., and LITTLE, W. A., 1953, *Proc. Phys. Soc.* A, **66**, 921.
BOWEN, E. J., and WILLIAMS, A. H., 1939, *Trans. Faraday Soc.*, **35**, 765.
KIRCHOFF, W., 1940, *Z. Phys.*, **116**, 115.
LIEBSON, S. H., BISHOP, M. E., and ELLIOTT, J. O., 1950, *Phys. Rev.*, **80**, 907.
MAERCKS, O., 1938, *Z. Phys.*, **109**, 685.
SWANK, R. K., and BUCK, W. L., 1955, *Rev. Sci. Instrum.*, **26**, 15.
WRIGHT, G. T., 1955, *Proc. Phys. Soc.* B, **68**, 929.

Point Contact Transistor Studies using Radioactive Collectors

BY D. HANEMAN AND A. J. MORTLOCK
Physics Research Laboratory, University of Reading

Communicated by R. W. Ditchburn; *MS. received* 1*st August* 1956, *and in revised form* 22*nd October* 1956

THE number of atoms transferred to a germanium base, whilst forming an antimony collector to produce enhanced current gain α in a point-contact transistor, has been measured in eleven cases. Pure Sb was used since it is a suitable material for forming (Haneman 1956). Also, the parameters associated with its diffusion in Ge have been previously studied. By combining these measurements with determinations of the surface diameters of the formed regions it has been possible to infer the existence of phase boundaries in these regions. Previous results for atom transfer from pulsed gold point contacts on Ge have been obtained by Aarons *et al.* (1954), but no transistor characteristics were recorded.

Pile-activated spectrographically pure Sb points containing the isotope ^{122}Sb were used in the present case, one to each 5 ohm cm single crystal of n-type Ge.

The activity transferred during forming was measured subsequently with a Geiger–Müller end-window counter. The surface distribution was found by autoradiography. Variable negative-voltage condenser discharges were used as forming pulses, and fine inactive tungsten points as emitters, the average spacing from the collector being $50\,\mu$. Forming was light and no signs of melting were apparent on the Ge surface when viewed under a magnification of ×20.

For eleven cases, the number of atoms transferred ranged from 2×10^{11} to 6×10^{14}: the values of current gain after forming ranged from less than 0·1 to 2·7. In only three out of nine cases was a single photographic spot image obtained in the expected position. In other cases considerable stray activity was transferred: the autoradiographs showed additional faint smears and, in some cases, multiple spots. These additional images are probably due to activity transferred during the small movements of the collector frequently necessary before a suitably rectifying point contact was established.

It was observed that simple low-pressure contact of an Sb point without pulsing also resulted in transfer of comparable amounts of material on the Ge. This could be removed with a very light etch whereas, with the formed specimens, a much heavier etch was required to remove about 95% of the total transferred activity. This indicated that the material transferred from simple contact was localized on the surface whereas, with formed specimens, it was in part continued to deeper levels. Results for the three single-image specimens in which the formed spot size could be measured unambiguously, are given in the table.

Specimen	Current Gain α Before	After	Diameter of Image	Total Number of Transferred Atoms
1	$<0{\cdot}1$	0·5	7μ	$0{\cdot}2 \times 10^{12}$
2	0·2	0·8	30μ	$7{\cdot}4 \times 10^{12}$
3	$<0{\cdot}1$	2·7	$\not> 300\mu$	$5{\cdot}0 \times 10^{12}$

Due to low activity per unit area the image size for specimen No. 3 was measured with relatively fast x-ray film, with a resolution considerably lower than that of the stripping film otherwise used; the diameter quoted can therefore be taken as an upper limit only. The figures given for the total number of transferred atoms can only be regarded as being accurate to a factor of about four, due to neglect of the effects of back-scattering, uncertainties in the effective geometry of the counter, etc.

It is clear that a complete range of quenched phases will exist in the region of the point of contact between the Sb and Ge after forming. However, it is not obvious what phases are retained by the affected Ge after removal of the collector. For the three specimens in the table one may calculate, using a simple diffusion geometry, a lower limit for the time required for location of the whole of the transferred atoms in the Ge-rich homogeneous α-phase of the Sb–Ge system, Using the figure given by Thurmond and Struthers (1953) for the maximum solubility of Sb in this phase at a temperature as high as 900°c (2×10^{-4} atom fraction), together with Dunlap's (1954) measured coefficient for α-phase diffusion at this temperature (2×10^{-10} cm^2 sec^{-1}), it is found that the times required are of the order of hours. Even if it is assumed that as little as 5% of the total number of transferred atoms penetrate below the Ge surface in the case of the three specimens noted in the table, this still applies.

Since the times during which the Ge is sufficiently hot for appreciable diffusion to take place are expected to be less than one second, it can be inferred that only a small proportion of the transferred atoms are in the α-phase. The remainder will be in part contained in other quenched phases between the α-phase zone and the Ge surface.

Enhanced hole trapping in the region of the formed collector has been suggested as a contributory factor in the production of high values of current multiplication after forming. It is noteworthy that phase boundaries in the affected region could act as a source of charge carrier traps. It is thus possible that the creation of such boundaries is a necessary process for successful forming.

ACKNOWLEDGMENTS

The authors wish to thank Dr. D. H. Tomlin for valuable advice and encouragement and have benefited from discussions with Dr. E. W. J. Mitchell. Acknowledgment is made to the Admiralty for permission to publish this note.

REFERENCES

AARONS, M. W., POBERESKIN, M., GATES, J. E., and DALE, E. B., 1954, *Phys. Rev.*, **95**, 1345.
DUNLAP, W. C., Jr., 1954, *Phys. Rev.*, **94**, 1531.
HANEMAN, D., 1956, *Proc. Phys. Soc.* B, **69**, 712.
THURMOND, C. D., and STRUTHERS, J. D., 1953, *J. Phys. Chem.*, **57**, 831.

The Absorption Spectra of Single Crystals of Pure and Partially Decomposed Silver Azide

BY A. C. McLAREN

Research Laboratory on the Physics and Chemistry of Surfaces,
The Department of Physical Chemistry, University of Cambridge

MS. received 17th *July* 1956, *and in revised form* 17th *October* 1956

As part of a general programme of research on the initiation and growth of reaction in the solid state, a study of the slow and explosive decomposition of the metallic azides is being made in this Laboratory. The measurement of the absorption spectra of these compounds is difficult since it is often impossible to grow single crystals either large enough or sufficiently perfect to be examined in commercial spectrophotometers. To overcome this difficulty a simple microspectrophotometer has been built whereby the visible absorption spectrum of a very small crystal, or of a small near-perfect area of an otherwise greatly imperfect crystal, can be measured. Magnifications up to × 400 have been used.

The apparatus is essentially a monochromator followed by a microscope and a radiation detector. Monochromatic light (obtained from a Siemens 375 watt xenon arc in conjunction with a Hilger quartz monochromator) was used to illuminate an ordinary optical microscope on which was mounted a fixed length camera with right-angle viewing eyepiece. The photographic plate holder was replaced by a photomultiplier tube and a sensitive galvanometer used to measure the current flowing. Provision was made for measurements with polarized light.

The absorption spectra were measured by taking readings at each wavelength of the transmission T through the selected portion of the crystal and the optics of the microscope, and of the transmission T_0 through the optics of the microscope only.

The absorption spectra of undecomposed AgN_3 for polarized light which is travelling in the [010] direction and whose electric vector is parallel to A, the [001] and B, the [100] directions are shown in figure 1. For brevity, absorption spectra corresponding to these directions will now be denoted by Spectrum A and Spectrum B respectively. The corresponding refractive indices have been measured by Evans and Yoffe (1957), the values being 1·82 and 2·20 respectively at 6500 Å.

An AgN_3 crystal was partially decomposed by the ultra-violet light from a 125 watt high pressure mercury lamp filtered with a Chance OX7 glass filter and 4 cm of water. The irradiation was continued for 2¾ hours with the crystal approximately 50 cm from the lamp. This corresponds to an intensity incident on the crystal of the order of 10^{14} quanta cm^{-2} sec^{-1} of wavelengths from 2500 to 4400 Å. In figure 2 it will be seen that this type of decomposition gives rise to a broad absorption band which overlaps the fundamental absorption edge and peaks at approximately 4000 Å in both spectra.

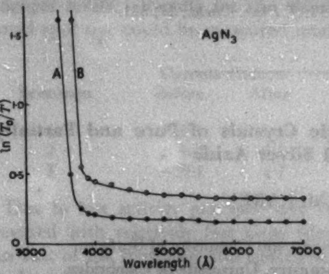


Figure 1. The absorption spectra of an undecomposed single crystal of silver azide.

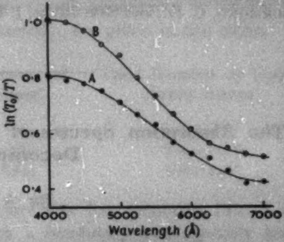


Figure 2. The absorption spectra of silver azide crystal which has been partially decomposed by irradiation with ultra-violet light.

The absorption spectra resulting from thermal decomposition at 115°c for 50 minutes are shown in figure 3. It will be seen that in spectrum A there is a new absorption band with peak at about 4750 Å, while in spectrum B the new band is displaced slightly to longer wavelengths with peak at about 4900 Å, An attempt has been made to interpret these results by a direct application of the general theory of light scattering by spherical conducting particles in a dielectric medium, which has been given by Mie (1908).

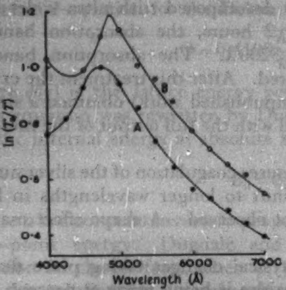


Figure 3. The absorption spectra of a silver azide crystal which has been partially decomposed by heating.

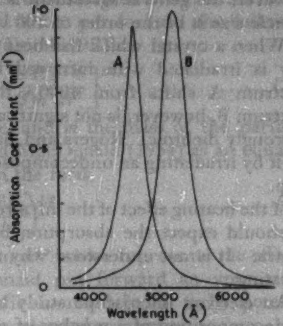


Figure 4. The absorption curves calculated from the Mie theory for silver azide containing one part per million by volume of metallic silver, for the limiting case in which the particles are much smaller than the wavelength of light.

With the optical constants of metallic silver as given by Minor (1903), the absorption curves for metallic silver spheres embedded in silver azide, for the limiting condition in which the particles are much smaller than the wavelength of light, have been calculated for the refractive indices 1·82 and 2·20. In figure 4 it will be seen that the peaks of the absorption curves are at 4650 Å and 5200 Å respectively.

The agreement with the absorption spectra as observed in crystals decomposed thermally is therefore quite good. On the other hand, the agreement is poor for the absorption bands resulting from ultra-violet irradiation. The difficulty can, however, be resolved in the following way. Since ultra-violet light is absorbed in a very thin surface layer of silver azide, all the decomposition must take place within this region. After considerable decomposition has taken place, it may be expected that the nuclei will tend to project from the surface of the crystal. According to the Mie theory, the absorption band resulting from scattering by silver spheres in air will have its peak at about 3650 Å. In practice, some of the nuclei may be expected to be embedded in the parent crystal, and this will tend to broaden the band and shift the peak to longer wavelengths.

These results are summarized as follows:

Mode of Decomposition	n_{eff}	λ (Å)	
		Theory	Expt.
Ultra-violet irradiation	1·00	3650	4000
Thermal	1·82	4650	4750
Thermal	2·20	5200	4900

n_{eff} = effective refractive index; λ = wavelength of peak absorption.

It will be noticed that the theoretical absorption bands are considerably sharper than those actually observed. This is probably due to the fact that in the theory we are dealing with ideal spherical particles all of the same size, while in practice deviations from ideal conditions will almost certainly take place. However, the general agreement is fairly good, and the theory suggests that the particle size is of the order of 100 Å.

When a crystal which has been partially decomposed with ultra-violet light only is irradiated with infra-red light for 2 hours, the absorption band in spectrum A shifts from 4000 Å to about 5200 Å. The absorption band in spectrum B, however, is not significantly altered. After this treatment the crystal is strongly dichroic. Rogers and Sawkill (unpublished work) obtained a similar result by irradiating an undecomposed crystal with the full output of the mercury lamp.

If the heating effect of the infra-red light causes coagulation of the silver nuclei, we should expect the absorption band to shift to longer wavelengths in both spectra. It is not understood why this is not observed. A shape effect may be operating.

An electron microscope study of the physical changes taking place during the decomposition of a number of metallic azides is being made at present, and it is hoped that this will give more information regarding the shape and size of the metallic nuclei.

ACKNOWLEDGMENTS

I thank Dr. F. P. Bowden for helpful discussions throughout this work, and Dr. J. Sawkill for his cooperation in making the calculations from the Mie theory, and for many discussions. I also thank Imperial Chemical Industries Limited for providing apparatus.

REFERENCES

EVANS, B. L., and YOFFE, A. D., 1957, *Proc. Roy. Soc.* A, **238**, 568.
MIE, G., 1908, *Ann. Phys., Lpz.*, **25**, 377.
MINOR, R. S., 1903, *Ann. Phys., Lpz.*, **10**, 581.

Melting Curves of Helium and Hydrogen Isotopes

BY C. DOMB

Wheatstone Laboratory, King's College, London

MS. received 16*th October* 1956

MILLS AND GRILLY (1955) have recently determined the melting curves of ^{4}He and ^{3}He over a considerable pressure range. For ^{4}He their results agreed with previous measurements of Simon and his collaborators (Holland *et al.* 1951, Dugdale and Simon 1953) and on fitting the data to a Simon melting curve

$$P=\alpha+\beta T^{c}, \qquad \ldots\ldots(1)$$

the value of α is negative, corresponding to a positive 'internal pressure'. For ^{3}He, on the other hand, the value of α is positive and this corresponds to a negative internal pressure. It is of interest to see how this result fits in with the theoretical treatment of solid helium.

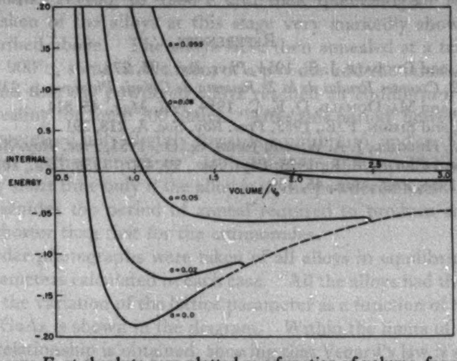
In an attempt to account for the properties of substances in which the role of zero-point energy is important, the present writer put forward the following differential equation for the Debye Θ value (Domb 1952):

$$\Theta = 0{\cdot}4109\frac{h}{k}\left[\frac{\phi''(r)}{m} + \frac{9}{8}\frac{R}{m}\frac{d^2\Theta}{dr^2}\right]^{1/2} \qquad \ldots\ldots(2)$$

where $\phi(r)$ is the lattice energy per particle, and m the mass of the particles. This equation was rewritten by Dugdale and MacDonald (1954) as an equation for the internal energy at absolute zero E_0 in the form

$$E_0 = \phi(r) + a\left(\frac{d^2E_0}{dr^2}\right)^{1/2} \qquad \ldots\ldots(3)$$

where a is a constant inversely proportional to $m^{1/2}$ which represents the effect of zero-point energy. Dugdale and MacDonald put forward an approximate solution in closed form when $\phi(r)$ is a Mie–Lennard-Jones 6–10 potential, and this solution has a number of interesting physical properties.



The internal energy E_0 at absolute zero plotted as a function of volume, for various values of a (see equation (3)).

The form of E_0 as a function of volume for varying a is shown in the figure. As a increases the minimum is pushed out and the potential well becomes shallower until, for a particular critical value a_c, E_0 ceases to have a minimum. It seems that ^{4}He with a positive internal pressure corresponds to a value of $a < a_c$, whereas ^{3}He corresponds to $a > a_c$. For ^{4}He the solid phase could exist under atmospheric pressure, but the liquid phase is more stable; for ^{3}He, however, the solid phase could not exist.

The accuracy of the solution of Dugdale and MacDonald in the region of applicability to ^{4}He and ^{3}He, and the dependence of the solution on the form of intermolecular potential are being investigated by M. E. Fisher and I. J. Zucker at King's College.

By means of measurements of the Debye Θ-values of ^{4}He along the melting curve Dugdale and Simon (1953) were able to show that the degree of degeneracy steadily increased as the melting temperatures and pressures increased, and there are also theoretical reasons for expecting this type of behaviour (Salter 1954). One might, therefore, expect that sufficiently far along the melting curve classical theory is valid, and the curves for ^{4}He and ^{3}He become identical. It is difficult to assess how much reliance can be placed on extrapolations based on the formulae of Mills and Grilly (1955), but it is interesting to note that despite the initial apparent parallelism of the melting curves, these formulae do lead to an intersection at a temperature of 55° and a pressure of 9000 atmospheres.

For the hydrogen isotopes H_2, D_2 and T_2 the structure of the molecule is to some extent dependent on the isotopic mass, and it is, therefore, no longer rigorously true to say that in the region where melting is classical the melting curves would be identical (except at much higher pressures where the substance becomes monatomic). Nevertheless, it is worth pointing out that despite the initial parallelism noted by Chester and Dugdale (1954) subsequent measurements by Mills and Grilly (1956) indicate that the melting curves are approaching one another at higher temperatures and pressures.

The writer is grateful to the late Sir Francis Simon for helpful discussions.

References

Chester, P. F., and Dugdale, J. S., 1954, *Phys. Rev.*, **95**, 278.
Domb, C., 1952, *Comptes Rendus de la 2^e Reunion de Chimie Physique*, p. 338.
Dugdale, J. S., and MacDonald, D. K. C., 1954, *Phil. Mag.*, **45**, 811.
Dugdale, J. S., and Simon, F. E., 1953, *Proc. Roy. Soc.* A, **218**, 291.
Holland, F. A., Huggill, J. A. W., and Jones, G. O., 1951, *Proc. Roy. Soc.* A, **207**, 268.
Mills, R. L., and Grilly, E. R., 1955, *Phys. Rev.*, **99**, 480; 1956, *Ibid.*, **101**, 1246.
Salter, L. S., 1954, *Phil. Mag.*, **45**, 369.

LETTERS TO THE EDITOR

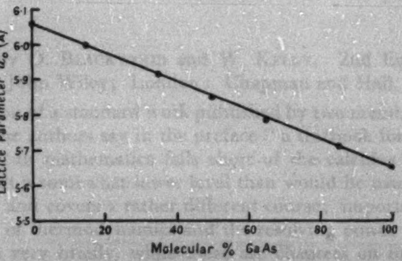
Solid Solution in the GaAs–InAs System

In a recent note (Woolley, Smith and Lees 1956) it has been shown that in the quasi-binary system GaSb–InSb, continuous solid solution occurs over the whole range of composition, but that this equilibrium condition is only established after annealing for times of the order 1500 hours at temperatures just below the solidus value. It was found that for all alloys investigated, with shorter times of anneal the x-ray powder photographs showed bands rather than sharp lines, indicating a non-equilibrium condition in which a large number of phases of varying composition occurred. As the period of anneal was increased, the bands slowly narrowed into lines, showing that the equilibrium condition was being established.

Similar results have now been obtained in the quasi-binary system GaAs–InAs; once again the equilibrium condition of solid solution throughout the whole range of composition was established only after a considerable time of anneal. Alloys containing 15, 35, 50, 65 and 85 molecular percentage GaAs were produced by sealing the appropriate amounts of GaAs and InAs under vacuum in quartz, heating to 1300°c and then quenching in water. X-ray photographs taken of the alloys at this stage very markedly showed the band structure described above. The alloys were then annealed at a temperature of approximately 900°c, some in the form of a powder and the others still in bulk.

It was found that the powdered alloys reached an equilibrium single phase state after annealing for some 300 hours. After this period, however, the alloys in bulk had progressed only slightly towards equilibrium, and so these were also powdered before further annealing. It thus appears that equilibrium is attained in a reasonably short time only if the alloy is in the powdered form. Also, in the case of the arsenides the period of anneal required to produce equilibrium is considerably shorter than that for the antimonides.

X-ray powder photographs were taken of all alloys in equilibrium condition and lattice parameters calculated in each case. All the alloys had the zinc blende structure, and the variation of the lattice parameter as a function of the molecular percentage of GaAs is shown in the diagram. Within the limits of experimental error a linear relationship is obtained, showing that Vegard's law is satisfied.



Variation of lattice parameter as a function of composition.

Further work is being carried out with this system including the determination of the complete phase diagram and the investigation, by infra-red methods, of the energy gap as a function of composition.

The authors are indebted to Professor L. F. Bates for the facilities of his laboratory, to Mr. J. T. Edmond of the Services Electronics Research Laboratory, Baldock, for provision of the InAs and GaAs used in the work, and to the Chief Scientist, Admiralty, for sponsoring the work.

Department of Physics, J. C. Woolley.
University of Nottingham. B. A. Smith.
13th November 1956.

Woolley, J. C., Smith, B. A., and Lees, D. G., 1956, *Proc. Phys. Soc.* B, **69**, 1339.

CORRIGENDUM

Defects with Several Trapping Levels in Semiconductors, by P. T. Landsberg (*Proc. Phys. Soc.* B, 1956, **69**, 1056).

On p. 1058, the phrase in brackets at the end of § (C) should have been deleted.

REVIEWS OF BOOKS

Propagation des ondes dans les milieux periodiques, by L. **Brillouin** and M. **Parodi**. Pp. 347. (Paris: Masson et Cie, 1956). Paper bound, 4000 fr; Cloth bound 4600 fr.

This monograph is an extended version of a book, published in the United States by Professor Brillouin, which will be familiar to many physicists.

About a third of the volume is concerned with the vibrations of one-dimensional lattices, both free and forced. The mathematically analogous case of loaded lines in electrical theory is also fully discussed. This is followed by an analysis of the propagation of waves in a static lattice in two and in three dimensions, where the lattice is considered to provide a small perturbation. A detailed discussion is given of reciprocal space and of Brillouin zones. The general case where the periodic field of the lattice is not small is treated in the one dimensional case using the Mathieu equation and the Hill determinant. The final part of the book deals with propagation of waves along four terminal lines and along wave guides.

The treatment of the material is in many respects original, one might almost say idiosyncratic. The historical account of the theory of the vibrations of a linear lattice is particularly welcome; but one misses any reference to the dynamics of two or three dimensional lattices, historical or otherwise. It is true that the Debye theory is mentioned and the vibrational spectrum of a crystal is derived, but this is obtained as a sum of the Debye spectra. That this is a very misleading representation has been clear for nearly twenty years, and it is regrettable that none of the dozen or so known vibrational spectra was reproduced. The references to the literature, while certainly adequate in the main are peculiarly deficient on the vibrational side. Though the mathematical methods introduced by Born and von Kármán are used, the relevant paper is not quoted. It also seems strange to find a discussion (however brief) of the expansion coefficient of crystals without any mention of the work of Grüneisen. On the other hand, the literature pertaining to Brillouin zones is quoted very fully.

These criticisms do not, however, detract from the value to a student of physics of a book which is interesting and stimulating. Within the limits set by the authors, it is exhaustive and rigorous, and can be strongly recommended.

M. BLACKMAN.

General Physics, by O. **Blackwood** and W. **Kelly**. 2nd Edn. Pp. x + 704. (New York: John Wiley; London: Chapman and Hall, 1955.) 54*s*.

This new edition of a standard work published by two members of Pittsburgh University is, as the authors say in the preface " a textbook for college students whose preparation in mathematics falls short of the calculus ". The book is therefore written at a somewhat lower level than would be usual for an English first-year student, and covers a rather different course; important subjects such as the second law of thermodynamics and the resolving power of optical instruments are treated very briefly, while there are chapters on meteorology, electronics, and nuclear physics.

Discussions and explanations are clearly and carefully written, the book is well illustrated, and each chapter is provided with a summary and a very full set of problems. It will be useful to first year undergraduates working for a physics degree, who will be stimulated by both the familiar and the unfamiliar parts of the text.

A. V. COHEN.

Principles and Applications of Physics, by O. Bluh and J. D. Elder. Pp. xiv + 866. (Edinburgh: Oliver and Boyd, 1955.) 45*s*.

The authors have written an elementary textbook of physics at quite a low level of discussion (say that of a University Entrance Scholarship) but addressed to a much more mature pupil than usually studies at that stage. Thus there are sections explaining the use of graph-paper and the action of the electric bell, and there are also sections discussing the nature of causality and neo-positivism, and the organizational problems of modern research laboratories.

The book will be valuable to two classes of readers. Students of subjects such as medicine or engineering who need some knowledge of physics will no doubt find this book very useful. The chapters on the philosophical aspects of physics may well interest a wider public, though a verbal discussion of the principle of uncertainty without a rigorous mathematical treatment is inevitably a little difficult if not confusing.

The final chapters contain a useful list of physics institutions, learned societies and journals, and some observations on the relation of physics to general scholarship.

A. V. COHEN.

Proceedings of the International Conference of Theoretical Physics, 1953, *Kyoto and Tokyo*, edited by I. Imai. Pp. xxviii + 942. (Tokyo: Science Council of Japan, 1954).

The published proceedings of the International Conference on theoretical physics, held in Kyoto and Tokyo in September 1953, form an enormous volume of some nine hundred and forty-two large pages. The contents cover most aspects of modern physics. Three hundred pages are devoted to Field Theory and Elementary Particles, Nuclear Physics takes up another seventy pages, Statistical Mechanics one hundred and eighty, Molecules and Solids three hundred and ten, and Liquid Helium and Superconductivity the remaining space. The name index of those taking part in the conference contains one hundred and sixty names, and includes many of the world's leading authorities in each of the above-mentioned fields. In addition to the printed version of the lectures there are many pages devoted to reporting the discussions which took place at each session. Both the lectures and the discussions were tape-recorded, and the reports have been constructed from these records with the minimum of modification.

Facts such as these ensure that the volume under review is an important addition to the literature of theoretical physics. Perhaps its most valuable aspect, to those not privileged to attend the meetings, is the opportunity which it gives to the reader of learning the views of the experts on unsettled and still controversial topics. Apart from this and the valuable synoptic view which the collection of articles provide, many readers may feel disappointment at their inability to understand quite a large fraction of the whole work. Many of the lectures are summaries, or reviews, of matter published in greater detail elsewhere and are readily intelligible only to those already well acquainted with such work. Nevertheless much of value may be gleaned by the diligent reader, and the organizers of the conference are to be congratulated on the production of this report.

As this was a truly international conference it is, perhaps, necessary to mention that, with only one exception, all papers are in English, or at least in that very nice international language for scientists: broken English. Even the articles by authors for whom English is the mother tongue are refreshingly colloquial in style as a result of the direct transcription from the tape-recordings.

H. JONES.

Theorie der Wärme, by R. BECKER. Pp. viii + 320. (Berlin: Springer, 1955.) DM. 39.60.

As would be expected by those familiar with the late Professor Becker's earlier writings, this book is an excellent modern introduction for physicists to the theory of heat. In the first half of the book the author expounds the basic facts of thermodynamics and of classical and quantum statistics. The material is of course fairly standardized but the author has made a serious attempt to link the statistics to the mechanics by means of the H-theorem and the properties of phase volume. Partly for this reason his explanation of the various ensembles used in statistical mechanics is delightfully lucid and compact. A minor achievement worth noting is in his clear discussion of the various magnetic energies which are introduced into thermodynamics.

The applications which are made in the second half of the book include: the theory of real and ideal gases (including the remarkable analogy between the real gas and the Bose gas); crystal vibrations and the order–disorder problems in solids; fluctuations and Brownian movement (including electrical noise); the thermodynamics of irreversibility. It is inevitable that such a short treatment of an enormous range of subjects will leave the experts unsatisfied. However the references are up-to-date and the book will be found by post-graduate students to be a useful introduction to the theory of heat. Most of the book is somewhat too difficult for the average English undergraduate student of physics.

R. O. DAVIES.

Introduction à la mécanique des milieux continus, par P. ANGLÈS D'AURIAC. Pp. 139. (Paris: Publications Scientifiques et Techniques de l'Air, No. 306, 1955.) 1300 fr.

This volume consists of a study of the mechanics of deformable media in terms of matrix algebra. The treatment is clear and the book is written in a style which lies in the best French tradition. It can be recommended as an introduction to the subject for those whose knowledge of mathematics is more profound than their acquaintance with physics or engineering. The topics treated in the book include the kinematics of deformation, the general analysis of stress and the general equations of equilibrium and of motion of fluids and elastic solids.

R.M.D.

A Textbook of Sound, by A. B. Wood, 3rd Edn. Pp. xvi + app. 638. (London: Bell, 1955). 42*s*.

This new edition of Dr. Wood's popular book follows the two previous editions of 1941 and 1930, each of which was reprinted several times. The author has brought this edition up to date by additional references and by about 60 extra pages of new material. At the same time nearly all the original contents of the book are preserved. By adopting the ingenious scheme of adding the extra pages two at a time and numbering them *n*A and *n*B, the printers have in most cases avoided having to reset subsequent page numbering.

Section I on the theory of vibrations is unaltered apart from the addition of references. Section II is amplified by extra pages on the behaviour of sound waves in tubes, absorption of sound in liquids by gas bubbles, edge tones and vortex systems, and on cavitation. Section III now includes discussions of phase and group velocity and propagation of modes in tubes, and extra material on transmission through plates and in the sea. The pages on transmission of sound in solids and liquids have been amplified and brought up to date in the light of the considerable amount of post war work.

In Section IV theories of hearing are treated even more fully than before as is also the theory of the Rayleigh disc in both gases and liquids. A short discussion of the reciprocity method of calibrating microphones and loudspeakers is introduced. Descriptions are also added of the sound spectrometer and the Debye effect in electrolytes.

The most extensive alterations have been made, as would be expected, to Section V on Technical Applications. This is now 17 pages longer than before and includes a new chapter on applications of ultrasonics. Acoustics of buildings and sound transmission through walls and floors are treated at greater length, and there are references to electronic musical instruments, sound reinforcement and a number of other new topics.

Without being entirely rewritten there must by now be a few things in so large a book that may strike younger physicists as old fashioned, the cathode ray oscillograph in Fig. 134 for example. Nevertheless, Dr. Wood's book is fully up to date and this edition seems certain to be at least as popular as the earlier ones.

J.M.M.P.

Physical Techniques in Biological Research; Vol. I. *Optical Techniques*. Edited by G. Oster and A. W. Pollister. Pp. xiii + 564. (New York and London: Acadamic Press, 1955) $13.50.

This volume is the first of three, the object of which is to collect together brief reviews of international authorship describing the main physical techniques at present used in biological research. We are told (preface) that the present volume (like Vol. II) deals with " methods applied to relatively pure biological substances ". Chapters on photochemistry, light scattering and absorption techniques are therefore included, although chapters also appear on the visible light microscope in its various forms (conventional, phase and interference) which are undoubtedly used to study the most impure biological materials. The editors have thus had some difficulty in following their original plan. The final chapters cover birefringence and dichroism, and electron microscopy.

Most chapters are written along similar lines, beginning with the theory of principles of the method and followed by a description of techniques and instrumentation illustrated with practical examples. In the first, entitled " Photochemistry and Luminescence " J. L. Rosenberg gives an account of the various internal energy transitions which can occur within a molecule which has absorbed radiation. The factors which determine any resulting chemical reaction are discussed and the chapter ends with a section on components used in experimental apparatus.

The next review (by G. Oster) gives a clear and concise theory of light scattering as applied to various particles, and a description of one of the American built light scattering photometers at present available.

Chapters 3, 4 and 5 deal with absorption spectrophotometry as applied to substances in macroscopic amounts. In the first of these C. F. Hiskey discusses the fundamental principles of spectroscopy, while in the succeeding chapter J. F. Scott deals with the application of these principles to ultraviolet spectrophotometry. The longest chapter in this book is that by Carl Clark on infrared spectrophotometry, and includes a full discussion and comparison of the various components and techniques used. The author also discusses the cataloguing of infra-red spectra as an aid to the interpretation of new data, and expresses once more the long felt need for some form of international catalogue of spectra. As in the preceding chapter, a comprehensive list of references is included.

The review of the light microscope by L. C. Martin provides a clearly written introduction to the subject, which is followed in Chapter 7 by a treatment of phase and interference microscopy by Osterberg. Unfortunately the description of the adjustment of the Dyson interference microscope in the latter chapter is, to say the least, misleading.

After a further short review by G. Oster on the principles and methods of birefringence and dichroism, the volume ends with a chapter on electron microscopy by V. E. Cosslett. This contains a useful and detailed review of specimen preparation techniques.

Any method of covering such a large and diverse subject as these volumes have attempted involves some disadvantage. The method adopted here has resulted in a certain amount of repetition; for instance, the resolving power of a microscope is discussed in two separate chapters. The work is also not entirely free from misprint. We are surprised to learn for example (page 447) that the difference in the refractive indices of a birefringent material is at all times equal to zero. There is no doubt, however, that the collection of these techniques in three volumes will provide a valuable source of reference. E. M. DEELEY.

CONTENTS OF SECTION A

PAGE

Dr. A. Sugie, Dr. P. E. Hodgson and Dr. H. H. Robertson. The Contribution of Tensor Forces to n–α Scattering 1

Dr. P. Baxter and Dr. F. R. Stannard. The Use of Multiple Scattering and the Effects of Distortion in a High-Pressure Cloud Chamber 17

Prof. J. D. Jackson. On the Use of the Complete Interaction Hamiltonian in Atomic Rearrangement Collisions 26

Dr. R. F. Barrow, Mr. P. G. Dodsworth and Prof. P. B. Zeeman. Rotational Analysis of Bands of the A $^3\Pi_0{}^+$, B $^3\Pi_1$—X $^1\Sigma^+$ Systems of Gallium Monofluoride 34

Dr. R. McWeeny and Mr. T. E. Peacock. The Electronic Structure and Spectra of some Nitrogen Heterobenzenes 41

Dr. D. B. Gayther and Mr. K. P. Nicholson. The Average Neutron Total Cross Sections of Heavy Elements at Low Energy 51

Dr. Mohamed El Nadi. Some Aspects of the Stripping Reactions 62

Mr. N. Sarma, Mr. M. Govindjee and Dr. H. R. Allan. A Study of the Reaction $^{12}C(d, p)^{13}C$ 68

Dr. J. M. Calvert, Dr. A. A. Jaffe and Mr. E. E. Maslin. An Investigation of the Reactions $^{12}C\ (d, n)^{13}N$ and $^{28}Si\ (d, n)^{29}P$ 78

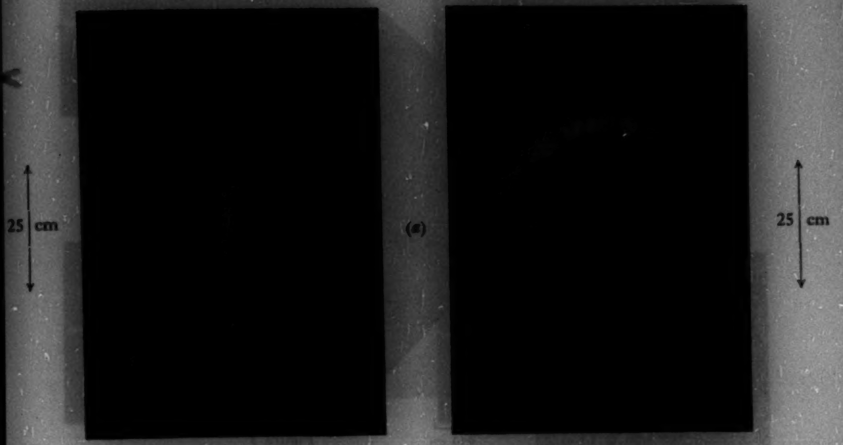
Mr. J. H. Towle, Mr. R. Berenbaum and Mr. J. H. Matthews. γ-Radiation following Proton Capture in ^{35}Cl, ^{37}Cl and ^{39}K 84

Letters to the Editor :

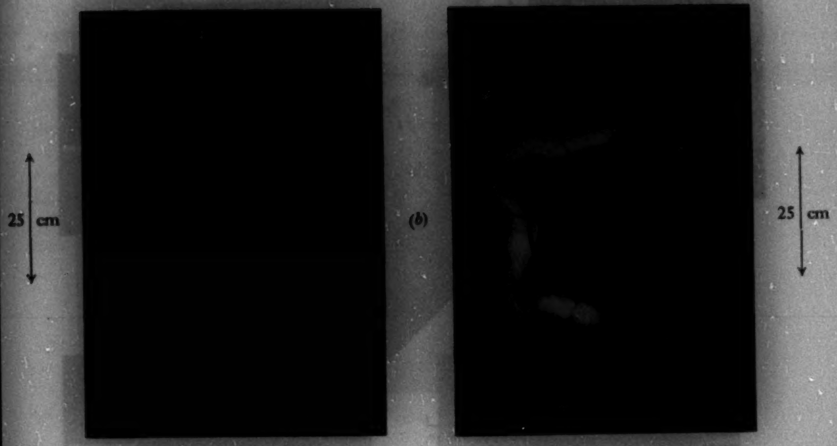
Mr. P. C. Russell and Prof. E. A. Stewardson A Re-measurement of the $L\alpha_{1,2}$ Lines of Lanthanum 91

Reviews of Books 93

Contents of Section B 96



(a)



(b)

Figure 2

Figure 5

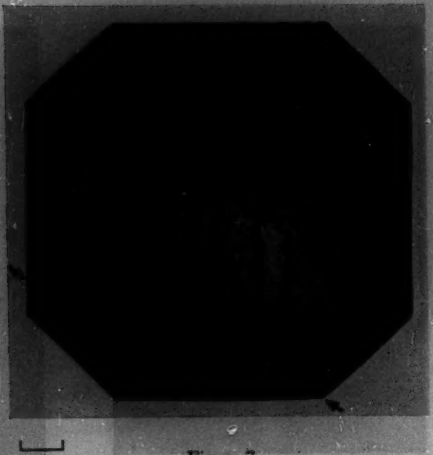


Figure 7

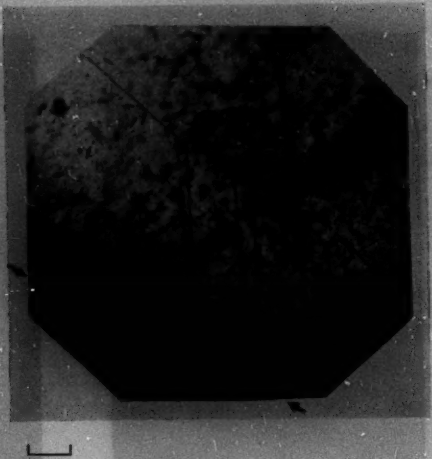


Figure 2. Powder pattern on a (111) surface.

Figure 5. Powder pattern showing echelon structure at the edge of crystal disc. (100) plane. Arrows mark the crystal's edge.

Figure 7. Powder pattern showing echelon structure at the edge (marked by arrows) of crystal disc. (100) plane.

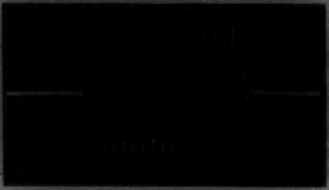


Figure 2. Electrical input.

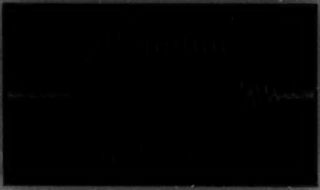


Figure 6. Mechanical pulse.

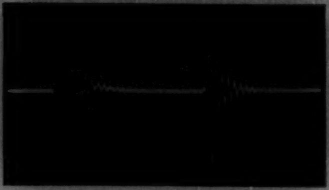


Figure 3. Reject at resonance.

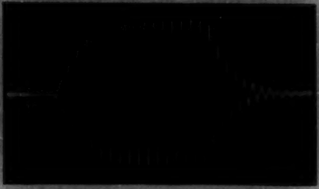


Figure 9. Pass at resonance.

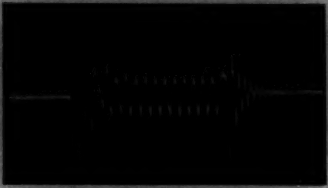


Figure 4. Reject near resonance.

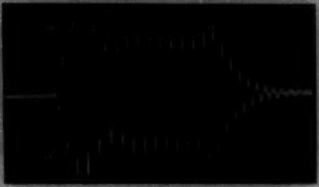


Figure 10. Pass near resonance.

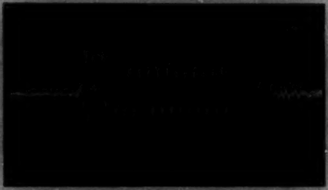


Figure 5. Transmitted pulse.

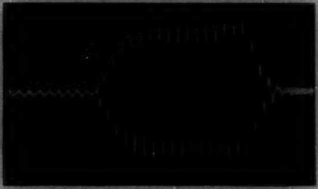


Figure 11. Reflected pulse.

Temperature Dependent Energy Levels in Statistical Mechanics

By E. W. ELCOCK and P. T. LANDSBERG

Department of Natural Philosophy, University of Aberdeen

MS. received 9th April 1956, *and in revised form* 25*th October* 1956

Abstract. A general statistical mechanical method for introducing temperature dependent 'energy levels' is developed. Previous discussions of this problem are included as particularizations of the general treatment and are unnecessarily restrictive. It is shown that appropriate definitions and procedures depend upon the nature of the information available or, alternatively, on the nature of the information it is desired to obtain. Simple and typical problems are considered in illustration of the general results.

---

§ 1. INTRODUCTION

IN an *ab initio* statistical mechanical treatment energy levels appear as purely mechanical concepts. As pointed out by Rushbrooke (1940), the use of explicitly 'temperature dependent energy levels' implies a process of prior averaging over some of the states of a system of the ensemble† under consideration. It does not seem to have been made clear, however, that, for a given system, different It kinds of prior averaging are possible depending on the nature of the information available or, alternatively, on the nature of the information it is desired to obtain. A lack of appreciation of this has led to insufficient flexibility in statements of procedures to be followed when explicitly temperature dependent energies are introduced into statistical mechanical treatments. This last statement will be amplified in the following sections. A new and general treatment of the problem of the introduction of temperature dependent energy levels into statistical mechanics is given in § 2. It is shown in § 3 how the procedure of Rushbrooke is included as a very special case of application of the general treatment to a particularly simple problem. Finally, in § 4, the general treatment is applied to a new class of problems, that is concerned with 'decomposable systems'.

§ 2. A GENERAL TREATMENT OF TEMPERATURE DEPENDENT ENERGY LEVELS

It is desired to prove the following result. Given a closed physical system the energy levels of which are specified by n symbols $i_1, i_2, \ldots i_n$, each of which represents one or a group of quantum numbers (or configuration variables), let $P_r(i_r)$ be the probability that the rth symbol has the value i_r. Let a strictly decreasing function $G(x)$ be given such that

$$P_r(i_r) = A_r G[\mathscr{E}_r(i_r)/\boldsymbol{k}T], \qquad \ldots\ldots(2.1)$$

where A_r is a dimensionless normalizing constant and the $\mathscr{E}_r(i_r)$ are quantities, with the dimensions of energy and in general temperature dependent, defined by (2.1). It is to be shown that, given the P_r and the function G, the $\mathscr{E}_r(i_r)$ can always be constructed (in general in an infinity of ways) so that any thermodynamic function for the whole system is a sum of m terms $(1 \leqslant m \leqslant n)$ each of which involves formally only one symbol i_r. By renumbering of symbols, if necessary, these m symbols can be taken to be $i_1, i_2 \dots i_m$.

---

† The word ensemble is used throughout this paper in the sense of a canonical ensemble: the arguments given are readily generalized to other ensembles.

Defining

$$\phi_r(i_r) \equiv \sum_{i_1,\dots i_{r-1},\, i_{r+1},\dots i_n} \exp\left[\frac{-E(i_1,\dots i_n)}{\boldsymbol{k}T}\right], \quad Z_r = \sum_{i_r} G\left[\frac{\mathscr{E}_r(i_r)}{\boldsymbol{k}T}\right], \qquad \dots\dots(2.2)$$

it follows from (2.1) that

$$P_r(i_r) = \phi_r(i_r)/Z = G[\mathscr{E}_r(i_r)/\boldsymbol{k}T]/Z_r, \qquad \dots\dots(2.3)$$

where Z is the usual partition function. That a set of $\mathscr{E}_r(i_r)$ satisfying (2.3) can be constructed in general in an infinity of ways follows immediately from the form (2.1) of (2.3): the different sets $\mathscr{E}_r(i_r)$ are given by different choices of the A_r. The restriction of $G(x)$ to be a strictly decreasing function of x is not necessary but ensures that the $P_r(i_r)$ decrease with an increase in the $\mathscr{E}_r(i_r)$ for given i_r, thus preserving the usual relation between energy and probability. Differentiating (2.3) with respect to the temperature and summing over all values of i_r, it is found that

$$\sum_{i_r}\left\{\frac{\partial \ln G[\mathscr{E}_r(i_r)/\boldsymbol{k}T]}{\partial T}\right\} P_r(i_r) = \frac{\partial \ln Z_r}{\partial T}. \qquad \dots\dots(2.4)$$

It is understood here that in all partial derivatives the volume of each system of the ensemble, and the number of systems in the ensemble, is kept constant. The convention is made that, after the $\mathscr{E}_r(i_r)$ have been defined for $r = 1, 2, \dots, m-1$, Z_m shall be defined by

$$\prod_{r=1}^{m} Z_r \equiv Z, \quad m \leqslant n, \qquad \dots\dots(2.5)$$

the levels $\mathscr{E}_m(i_m)$ being then again defined by (2.3). Summing (2.4) over all r from 1 to m, and writing U for the mean energy as averaged over the ensemble, it is found that

$$\left.\begin{aligned} U &= \sum_{r=1}^{m} U_r, \\ U_r &\equiv \boldsymbol{k}T^2 \sum_{i_r}\left\{\frac{\partial \ln G[\mathscr{E}_r(i_r)/\boldsymbol{k}T]}{\partial T}\right\} P_r(i_r). \end{aligned}\right\} \qquad \dots\dots(2.6)$$

This proves the required proposition for the mean energy. To prove it for the entropy it is sufficient to note that $S = U/T + \boldsymbol{k} \ln Z$, when it follows immediately from (2.5) and (2.6) that

$$\left.\begin{aligned} S &= \sum_{r=1}^{m} S_r, \\ S_r &\equiv \boldsymbol{k}T \sum_{i_r}\left\{\frac{\partial \ln G[\mathscr{E}_r(i_r)/\boldsymbol{k}T]}{\partial T}\right\} P_r(i_r) + \boldsymbol{k} \ln Z_r. \end{aligned}\right\} \qquad \dots\dots(2.7)$$

Using the relation $F = U - TS$ for the free energy it follows from (2.6) and (2.7) that

$$\left.\begin{aligned} F &= \sum_{r=1}^{m} F_r, \\ F_r &\equiv U_r - TS_r = -\boldsymbol{k}T \ln Z_r. \end{aligned}\right\} \qquad \dots\dots(2.8)$$

The quantities U_r, S_r, F_r satisfy the usual thermodynamic relations among each other as is readily verified. The proposition is therefore proved. It should also be noted that equilibrium mean values of quantities which depend only upon the state variables included in a particular symbol i_s, $1 \leqslant s \leqslant m$, are given by

$$\bar{x}(i_s) = \sum_{i_s} x(i_s) P_s(i_s), \qquad \ldots\ldots(2.9)$$

i.e. expressions which involve formally only one state symbol.

The expression (2.7) for the entropy can be put in the form

$$S_r = -\boldsymbol{k} \sum_{i_r} P_r(i_r) \left\{ \ln P_r(i_r) - \frac{\partial}{\partial T} (T \ln G[\mathscr{E}_r(i_r)/\boldsymbol{k}T]) \right\}, \qquad \ldots\ldots(2.10)$$

which shows the relation to the conventional expression for the entropy more clearly.

Regarding the $\mathscr{E}_r(i_r)$ as 'temperature dependent levels', a thermodynamic function will be said to be in its *standard form* when it is expressed in terms of the $\mathscr{E}_r(i_r)$ and $P_r(i_r)$. Then, while the expression (2.8) is a *usual* form for the free energy, the corresponding standard form

$$F_r = -\boldsymbol{k}T \ln \sum_{i_r} G[\mathscr{E}_r(i_r)/\boldsymbol{k}T]$$

is not a usual one unless

$$G(x) = e^{-x}. \qquad \ldots\ldots(2.11)$$

With this choice, it follows from (2.3) and (2.8) that

$$\left.\begin{aligned} &\mathscr{E}_r(i_r) - F_r = \overline{F}(i_r) - F, \\ \text{where}\quad &\overline{F}(i_r) \equiv -\boldsymbol{k}T \ln \sum_{i_1,\ldots i_{r-1}, i_{r+1},\ldots i_n} \exp[-E(i_1,\ldots i_n)/\boldsymbol{k}T], \end{aligned}\right\} \qquad \ldots\ldots(2.12)$$

and is the equilibrium free energy of the system inhibited to the state i_r. It is seen, incidentally, that here $\mathscr{E}_r(i_r)$ has the nature of a free energy. From (2.6) and (2.11),

$$\left.\begin{aligned} U_r &= \sum_{i_r} P_r(i_r) \left\{ \mathscr{E}_r(i_r) - T \frac{\partial \mathscr{E}_r(i_r)}{\partial T} \right\}, \\ S_r &= -\boldsymbol{k} \sum_{i_r} P_r(i_r) \left\{ \ln P_r(i_r) + \frac{\partial \mathscr{E}_r(i_r)}{\partial T} \right\}. \end{aligned}\right\} \qquad \ldots\ldots(2.13)$$

Both functions (2.13), in their standard forms, differ from the usual expressions by the incidence of the last terms. The omission of these terms from either or both of the thermodynamic functions (2.13) is in general incorrect and leads to violations of standard thermodynamic relations (Landsberg 1954).

If it is desired to have a usual standard form for U_r it follows from (2.6) that $G(x)$ must be chosen so that

$$\ln G[x(T)] = \int_{T_0}^{T} \left\{ \frac{x(y)}{y} \right\} dy, \qquad \ldots\ldots(2.14)$$

where T_0 is a fixed temperature. With this choice, it follows from (2.3) and (2.6) that

$$\left.\begin{aligned} &\mathscr{E}_r(i_r) - U_r = \overline{U}(i_r) - U, \\ \text{where}\quad &\overline{U}(i_r) \equiv \sum_{i_1,\ldots i_{r-1},\ldots i_{r+1},\ldots i_n} E(i_1,\ldots i_n) P(i_1,\ldots i_n) \Big/ \sum_{i_1\ldots i_{r-1}, i_{r+1}\ldots i_n} P(i_1,\ldots i_n), \end{aligned}\right\} \qquad \ldots\ldots(2.15)$$

and is the equilibrium internal energy of the system inhibited to the state i_r. Here $\mathcal{E}_r(i_r)$ has the nature of an internal energy. From (2.7) and (2.8),

$$\left.\begin{aligned} S_r &= -\boldsymbol{k}\sum_{i_r} P_r(i_r)\left\{\ln P_r(i_r) - \frac{\mathcal{E}_r(i_r)}{\boldsymbol{k}T} - \int_{T_0}^{T} \frac{\mathcal{E}_r(i_r)}{\boldsymbol{k}y^2}\,dy\right\}, \\ F_r &= -\boldsymbol{k}T\ln\sum_{i_r}\exp\left[\int_{T_0}^{T}\frac{\mathcal{E}_r(i_r)}{\boldsymbol{k}y^2}\,dy\right]. \end{aligned}\right\} \quad \ldots(2.16)$$

With this choice both functions (2.16), in their standard forms, differ from the usual expressions.

Neither of the choices (2.11) or (2.14) of the function G puts the expression for the entropy S_r in a usual standard form $S_r = -\boldsymbol{k}\ \Sigma_{i_r} P_r(i_r) \ln P_r(i_r)$. If this could be achieved for some choice of G then it would follow that

$$S = -\boldsymbol{k}\sum_r\sum_{i_r} P_r(i_r)\ln P_r(i_r),$$

and this is independent of the choice of G because the $P_r(i_r)$ are. In general, therefore, there is no function G which will achieve for the entropy what (2.11) and (2.14) achieve for the free energy and energy respectively.

There is clearly an infinity of choices of the function G for which *none* of the thermodynamic functions is in a usual standard form.

§ 3. EXACT EQUILIBRIUM PROPERTIES

It must be emphasized that the introduction of temperature dependent energy levels does not, in principle, imply an approximation: the thermodynamic functions given by (2.6), (2.7) and (2.8) are exact. The introduction of such levels may, in certain circumstances, considerably expedite the calculation of equilibrium mean values of interest. A particularly simple example is given by taking $m = 1$ in the general analysis of § 2.

Consider an ensemble A of systems of interest at temperature T. Let the states of a system of A be specified by the suffixes $i_1, i_2 \ldots i_n$ which, as in § 2, symbolize quantum numbers or groups of quantum numbers in quantum statistics, or configuration variables in classical or semi-classical statistics. Suppose that interest centres only in the state variables symbolized by i_1, and that it is desired to express information about equilibrium properties of A in terms of quantities which make explicit reference to i_1 only. It is seen from (2.6), (2.7), (2.8) and (2.9) that putting $m = 1$ in the general analysis of § 2 does just this. The choice (2.14) for G then leads to Rushbrooke's results (1940) and, in particular, from (2.9) readily follows Rushbrooke's statement that "... the correct procedure (for the evaluation of mean values of functions of variables not involved in the prior summations leading to the temperature dependence) when an energy level $\mathcal{E}$ depends on the temperature is to replace the Boltzmann factor $\exp[-\mathcal{E}/\boldsymbol{k}T]$, which would be involved were $\mathcal{E}$ constant, by $\exp[-\mathcal{F}/\boldsymbol{k}T]$ where $\mathcal{F}$ is the free energy associated with $\mathcal{E}$ by the Gibbs–Helmholtz relation $\mathcal{F} - T(\partial\mathcal{F}/\partial T)_{N,V} = \mathcal{E}$".† Rushbrooke (1940) has shown how the introduction of such temperature dependent levels may expedite the calculation of absorption isotherms. Other examples could be given. It should be emphasized here, however, that Rushbrooke's statement is valid only for $m = 1$ and the sometimes

† The interpolation in brackets is inserted by us to clarify the quotation.

convenient but arbitrary choice (2.14) of the function G. For this choice, as is seen from (2.15),

$$\mathscr{E}(i_1) = \sum_{i_2,\ldots i_n} E(i_1,\ldots i_n) P(i_1,\ldots i_n) \Big/ \sum_{i_2\ldots i_n} P(i_1,\ldots i_n), \qquad \ldots\ldots(3.1)$$

and is to be interpreted as the mean energy of a system of A, as averaged over all states of the system compatible with the given value of i_1. Other choices of the quantities $\mathscr{E}(i_1)$ are possible, and in fact frequently used in the literature, for which the statement about replacement of Boltzmann factors is no longer true: for example, in the treatment of regular assemblies (Fowler and Guggenheim 1939, §§ 610, 1313) it is often convenient to introduce quantities $\mathscr{E}(i_1)$ based on the choice (2.11) for G when $\mathscr{E}(i_1)$ has the nature of a free energy for the inhibited assembly. In the case of interest here ($m=1$), the choice (2.11) for G could be brought into correspondence with Rushbrooke's usage by calling the associated quantity $\mathscr{E}(i_1)$ given by (2.12) $\mathscr{F}(i_1)$: this $\mathscr{F}(i_1)$ and the $\mathscr{E}(i_1)$ given by (3.1) are related by the Gibbs–Helmholtz equation and the equivalence of the different procedures is immediately apparent. This change of notation will not be made, however, since the result it embodies is valid only for the special case $m=1$: it is neither necessary nor desirable in general to emphasize whether the quantities $\mathscr{E}_r(i_r)$ have the nature of a true energy (in the sense of the first law of thermodynamics) or a free energy. The defining equation (2.3) is quite general and allows choices of G, as and when convenient, for which the $\mathscr{E}_r(i_r)$ have neither the nature of a true energy nor a free energy.

As already mentioned, temperature dependent quantities $\mathscr{E}_1(i_1)$ are usually introduced for reasons of expedition and convenience, and their introduction does not imply any lack of knowledge about the complete set of energy levels of a system of the ensemble. Occasionally, however, procedures such as those generalized and systematized in § 2 may suggest profitable methods of approximation in situations in which there is such lack of knowledge (cf. from this point of view the discussion of the zeroth approximation for order–disorder in binary alloys (Fowler and Guggenheim 1939, § 1314)).

§ 4. EQUILIBRIUM PROPERTIES OF DECOMPOSABLE SYSTEMS

The procedures so far considered (§ 3) reflect the fact that interest centres on some only of the variables specifying the state of a system of the ensemble: it has nowhere been implied, however, that a system of the ensemble could be mentally decomposed into parts such that the variables of interest referred to one part only of the system. Such decomposable systems are of frequent interest and the general methods of § 2 are particularly well adapted to their investigation.

Consider, for example, an ensemble of systems A at temperature T. Let each system of A be mentally decomposable into two parts 1 and 2 in interaction. It is supposed that, to a first approximation neglecting the interaction, the state of the part system 1 may be specified solely by i_1 and the state of the part system 2 solely by i_2, where the symbolism (i_1, i_2) is as in § 2, when the energy of the part system 1 (2) in the state $i_1(i_2)$ may be written as $E_1(i_1)$ $(E_2(i_2))$. It is supposed that, to a good approximation, the interaction has the effect that the state of a complete system may still be specified by the two symbols i_1, i_2 but that the energy of the complete system in such a state can be written as

$$E(i_1, i_2) = E_1(i_1, i_2) + E_2(i_2). \qquad \ldots\ldots\ldots(4.1)$$

$E_1(i_1, i_2)$ may, for definiteness, be thought of as $E_1(i_1)+\epsilon(i_1, i_2)$ where $\epsilon(i_1, i_2) \ll E_1(i_1)$, but this is not necessary. The form of energy level structure (4.1) is introduced for definiteness of exposition: it is not necessary for the subsequent analysis which may be modified to take account of other level structures of interest. It is to be emphasized that the results of this section are exact once (4.1) is granted: the assumption involved in this equation is the only one required for this section.

Putting $m=2$ in the analysis of § 2, it is readily shown from (2.4) that the (dependent) equations (2.3) may be written as the single equation

$$G\left[\frac{\mathscr{E}_1(i_1)}{kT}\right] G\left[\frac{\mathscr{E}_2(i_2)}{kT}\right] = \left\{\sum_{i_2} \exp\left[-\frac{E(i_1, i_2)}{kT}\right]\right\}\left\{\sum_{i_1} \exp\left[-\frac{E(i_1, i_2)}{kT}\right]\right\} \Big/ \sum_{i_1 i_2} \exp\left[-\frac{E(i_1, i_2)}{kT}\right]. \qquad \ldots\ldots(4.2)$$

It is desirable that the temperature dependent levels $\mathscr{E}_1(i_1)$ and $\mathscr{E}_2(i_2)$ should, in the approximation in which the interaction between the part systems is neglected, reduce to $E_1(i_1)$ and $E_2(i_2)$ respectively. It is seen from (4.2) that this imposes the restriction

$$G\,[\mathscr{E}_j(i_j)/kT] \to \exp[-E_j(i_j)/kT] \text{ as } \mathscr{E}_j(i_j) \to E_j(i_j) \qquad (j=1, 2) \qquad \ldots\ldots(4.3)$$

on the function G. It is satisfied by the particular choices (2.11), (2.14) discussed previously.

It is intended that, when the thermodynamic functions are written as a sum of two terms (equations (2.6) to (2.8)), the thermodynamic functions labelled '2' and derived from $\mathscr{E}_2(i_2)$ should be those for an assembly of part systems with states i_2 and energies $E_2(i_2)$ considered as isolated, when it follows that the thermodynamic functions labelled '1' and derived from $\mathscr{E}_1(i_1)$ will be just those for the part system 1, which is taken to include the contribution due to the interaction between the part systems, but is otherwise regarded as isolated. This can be achieved by putting

$$\sum_{i_2} G[\mathscr{E}_2(i_2)/kT] = \sum_{i_2} \exp[-E_2(i_2)/kT]. \qquad \ldots\ldots(4.4)$$

Suppose G has the property that $y=G(x)$ implies a unique reciprocal relation

$$x=g(y), \qquad \ldots\ldots(4.5)$$

and conversely. Then, fixing G and Z_2 completely defines the $\mathscr{E}_1(i_1)$ and $\mathscr{E}_2(i_2)$ since, from (2.4) and (4.2),

$$\frac{\mathscr{E}_1(i_1)}{kT} = g\left\{\sum_{i_2} \exp\left[-\frac{E(i_1, i_2)}{kT}\right] \Big/ \sum_{i_2} \exp\left[-\frac{E_2(i_2)}{kT}\right]\right\},$$
$$\frac{\mathscr{E}_2(i_2)}{kT} = g\left\{\sum_{i_1} \exp\left[-\frac{E(i_1, i_2)}{kT}\right] \sum_{i_2} \exp\left[-\frac{E_2(i_2)}{kT}\right] \Big/ \sum_{i_1, i_2} \exp\left[-\frac{E(i_1, i_2)}{kT}\right]\right\}. \qquad \ldots\ldots(4.6)$$

It is readily seen that $\mathscr{E}_1(i_1)$ and $\mathscr{E}_2(i_2)$ do, in fact, reduce to $E_1(i_1)$ and $E_2(i_2)$ respectively in the approximation in which the interaction is neglected.

It is perhaps of incidental interest that, from (4.6), $\mathscr{E}_1(i_1)$ may be written as

$$\mathscr{E}_1(i_1) = kTg\left\{\sum_{i_2} \exp\left[-\frac{E_1(i_1, i_2)}{kT}\right] \exp\left[-\frac{E_2(i_2)}{kT}\right] \Big/ \sum_{i_2} \exp\left[-\frac{E_2(i_2)}{kT}\right]\right\}, \qquad \ldots\ldots(4.7)$$

when expanding the first exponential,

$$\mathscr{E}_1(i_1) = \boldsymbol{k}Tg\{1 - [E_1(i_1)]_{\text{av}}/\boldsymbol{k}T + [E_1^2(i_1)]_{\text{av}}/2!\,(\boldsymbol{k}T)^2 - \ldots\}, \qquad \ldots\ldots(4.8)$$

where the suffix av denotes what may be regarded as an average evaluated in the second part system,

$$[E_1^r(i_1)]_{\text{av}} \equiv \sum_{i_2} E_1^r(i_1, i_2) \exp[-E_2(i_2)/\boldsymbol{k}T] / \sum_{i_2} \exp[-E_2(i_2)/\boldsymbol{k}T]. \qquad \ldots(4.9)$$

It is seen that if (2.11) holds and if fluctuations in the second part system are neglected, then, to this approximation

$$\mathscr{E}_1(i_1) \simeq [E_1(i_1)]_{\text{av}}. \qquad \ldots\ldots(4.10)$$

This shows that Eyring's use of temperature dependent levels (Ewell and Eyring 1937, criticized by Rushbrooke 1940) can, in fact, be upheld (see below) as a first approximation.

The analysis just described is most useful when statistical mechanical knowledge of the system is incomplete. Many phenomena are initially treated as properties of an isolated system whereas, in fact, the system is in unknown interaction with some environment. The analysis above allows such an unknown interaction to be consistently, if empirically, treated whenever the system may be suitably decomposed. A concrete illustration may be taken from the molecular field treatment of the magnetic properties of crystals. In the first approximation a system of interacting spins on the sites of the crystal lattice is considered as an isolated system. The symbol i_1 in the analysis above is here just the relative magnetization of the spin system and the interaction between the spins is specified by a parameter, λ say (effectively an exchange integral), supposed given: the energy of the state i_1 of the spin system may then be written as $E_1(i_1, \lambda)$. A consistent account of the interaction of the spin system with the crystal environment may now be developed by considering the parameter λ as a function of the variables, i_2 say, specifying the state of the rest of the crystal system and approximating to the energy of the complete system by the expression

$$E(i_1, i_2) = E_1[i_1, \lambda(i_2)] + E_2(i_2), \qquad \ldots\ldots(4.11)$$

i.e. an expression of the form (4.1). The development may then be continued tentatively by assuming a particular form for the quantities $\mathscr{E}_1(i_1)$ given by (4.6), the choice being guided by the general result that $\mathscr{E}_1(i_1)$ reduces to $E_1(i_1, \lambda)$ in the approximation in which the interaction with the crystal environment is neglected. Since, without any necessity for detailed examination, the general analysis above tells us that the thermodynamic functions '2' are those for the crystal environment considered as isolated, we can now investigate, by means of equations (2.6) to (2.9) for the thermodynamic functions '1' alone, the effect on the magnetization and other equilibrium properties of the spin system of the postulated interaction between the spin system and the crystal environment. It has been shown (Elcock 1954) that such a development allows the prediction of (observed) transitions, which would not occur in an approximation which treats the spin system as isolated, between states of different magnetic structure.

The results of §§3 and 4 also have a bearing on the following problem. Consider formulae of the kind

$$x = A \exp(-\Delta E/\boldsymbol{k}T), \qquad \ldots\ldots(4.12)$$

which occur frequently in physics and physical chemistry. In this equation

ΔE represents the difference in energy of two quantum states of different i_1 say. The relation between log x and $1/T$ should, if ΔE were constant, be linear. Frequently the experimental results show a departure from linearity following a relation

$$x = A \exp[-f(T)/\boldsymbol{k}T], \qquad \ldots\ldots(4.13)$$

where $f(T)$ is a slowly varying function of T. The question arises as to the interpretation of $f(T)$. Now (4.12) usually expresses an approximate result derived by considering only a part of a given system (e.g. the electron gas in a semiconductor crystal). The lack of linearity found experimentally is usually taken as implying the inadequacy of this approximation and as resulting from the unconsidered interaction between the part system and its environment. In the earlier literature the perhaps naive view was expressed that $f(T)$ in (4.13) should be interpreted as an equilibrium mean value, $\Delta\mathscr{E}$ say, of ΔE at the temperature T, the mean being a suitable average over the states of the environment of the part system. The results of §4 show that, if an energy level structure such as (4.1) is supposed appropriate to the part system of interest and its environment, then, to the approximation involved in equation (4.10), this naive view is correct provided that $\Delta\mathscr{E}$ is interpreted as the difference of two mean values of the kind appearing on the right-hand side of (4.10). Rushbrooke (1940, 1949), on the basis of his result for the replacement of Boltzmann factors in a statistics making use of temperature dependent quantities suggests, on the other hand, that this view is wrong and that $f(T)$ is to be interpreted as

$$\boldsymbol{k}T\int_{T_0}^{T} \{\Delta\mathscr{E}(y)/\boldsymbol{k}y^2\}dy.$$

The results of §3 show, however, that this last expression is only correct if $\Delta\mathscr{E}$ is interpreted as a difference of the two mean values each of which is given by an expression of the form (3.1) where i_1 has different values for each of the two mean values. The results of §3 show also that $f(T)$ can be interpreted as the difference of two mean values each of which is of the form

$$\mathscr{E}(i_1) = -\boldsymbol{k}T\ln \sum_{i_2 \ldots i_n} \exp[-E(i_1, i_2 \ldots . i_n)/\boldsymbol{k}T],$$

where i_1 has different values for the two mean values. Other possibilities based on the analysis of §§2 and 3 exist.

ACKNOWLEDGMENT

We are indebted to Professor E. A. Guggenheim for correspondence at an early stage of this work.

REFERENCES

ELCOCK, E. W., 1954, *Phys. Rev.*, **94**, 1070.

EWELL, R. H., and EYRING, H., 1937, *J. Chem. Phys.*, **5**, 726.

FOWLER, R. H., and GUGGENHEIM, E. A., 1939, *Statistical Thermodynamics* (Cambridge: University Press).

LANDSBERG, P. T., 1954, *Phys. Rev.*, **95**, 643.

RUSHBROOKE, G. S., 1940, *Trans. Faraday Soc.*, **36**, 1055; 1949, *Introdnction to Statistical Mechanics* (Oxford: University Press), p. 316.

Directional Effects in the Electric Breakdown of Single Crystals of KCl and NaCl

By R. COOPER, D. T. GROSSART and A. A. WALLACE
Electrical Engineering Department, University of Manchester

MS. received 4th May 1956, and in revised form 17th October 1956

Abstract. In alkali halide crystals the discharge paths formed when electric breakdown occurs about a point electrode follow definite crystallographic directions, but many believe that the intrinsic electric strength of these materials does not depend on the direction of the applied field. This belief is difficult to reconcile with the above fact and examination of the literature shows that it is based on very scanty evidence.

The experiments described in this paper were made to extend the available information, and measurements of intrinsic strength were made with both KCl and NaCl between −200°c and +230°c. The results of the measurements provide little support for the above belief. A fair measure of agreement existed between the discharge track directions and the directions of smallest electric strength. This result suggests that the mechanism of breakdown is dependent on the properties of high-energy conduction electrons.

§ 1. Introduction

When an alkali halide crystal breaks down in the divergent electric field about a point electrode, the discharge tracks follow definite crystallographic directions. Since the electric field in this electrode system is spherically symmetrical, it seems that the intrinsic electric strength of the crystal is smaller in certain directions than it is in others. But, rather surprisingly, this view is not generally accepted and with respect to electric strength the alkali halides are considered by many to be isotropic.

The origin of this belief appears to be in a few measurements made by Davisson and von Hippel (1940) who reported that the electric strength of NaCl at room temperature was the same with the field applied in the (100) and the (110) directions, but was somewhat lower when the field was in a (111) direction. Davisson and von Hippel expressed doubt about the validity of this difference and they attributed it to the difficulty experienced in avoiding cracks in the (111) specimens, which they stated were mechanically the weakest.

Only tentative conclusions may be drawn from the results of Davisson and von Hippel but, in the absence of any other information, Callen and Offenbacher (1953) assumed that other alkali halide crystals, at all temperatures, would behave in the same way and they have proposed a theory of the discharge directions whose predictions are in reasonable agreement with the available observations.

Callen and Offenbacher assume that the scattering of conduction electrons is mainly due to interactions with the longitudinal optical modes of the lattice vibrations. The prediction of directed breakdown tracks depends upon an estimation of the scattering probability as a function of electron energy for different directions of electron motion in the crystal lattice. The scattering probability is shown to be independent of direction for electrons with energy less than about 4 ev, but for higher energies it depends on the Brillouin-zone geometry. The (111) direction has the greatest and the (100) direction has the smallest scattering probability for electrons of energies near to the ionizing energies but, for electrons of moderate energy, the corresponding directions are respectively (100) and (110). It is assumed that breakdown is due to an avalanche mechanism and electric strength is, therefore, the field in which electrons of about 1/10 ev energy can be accelerated up to the lattice ionization energy. During the period of acceleration the applied field tends to make the conduction electrons move in a direction parallel to it, but the lattice scattering tends to cause them to move in the crystallographic direction of least scattering. It is assumed that by far the greatest portion of the displacement of an electron between successive ionizing collisions is traversed at a particular energy called the dominant energy. The value of this energy determines the direction of least scattering probability and it is fixed intuitively by assuming it to be the highest energy at which the collision frequency is able to establish anisotropy against the action of the field. The dominant energy increases with rise in temperature and consequently, as the temperature rises, it is predicted that the breakdown track directions should follow the sequence

$$(\text{Random or } 100), \rightarrow (111), \rightarrow \begin{matrix}110\\ \text{or}\\ 100\end{matrix} \rightarrow (\ 100).$$

The breakdown tracks should take the (111) direction in KCl between about − 110°c and − 30°c and the same track direction should be observed between about − 50°c and + 15°c in NaCl.

Various theories of breakdown in crystalline solids have been proposed, but from the available experimental observations it is not yet possible to determine which one most likely represents the mechanism of breakdown in the alkali halides. An alternative to von Hippel's low-energy hypothesis is the proposal due to Fröhlich (1939) that only conduction electrons possessing energy near to the ionization energy of the lattice need be considered in determining when instability occurs, leading to disruption of the material. If the electric strength of a crystal is determined by the properties of conduction electrons with high energy it is reasonable to expect directional effects in the intrinsic electric strength of the material as well as the phenomenon of directed breakdown tracks. No calculations appear to have been made of the relative magnitudes of various values of the electric strength, although on the basis of Callen and Offenbacher's scattering probability calculation we may expect the electric strength to be least in the (100) direction and greatest in the (111) direction.

§ 2. Intrinsic Strength Measurements

The methods adopted for preparing specimens and measuring their electric strengths were very similar to those previously described by Calderwood, Cooper and Wallace (1953), who drew attention to the very great influence of internal strains on the electric strength of potassium chloride crystals. Internal strains may be produced in the alkali halide crystals when insufficient care is taken in their handling and no quantitative method exists of allowing for the effect on electric strength of the associated deformation of the crystal lattice. Consequently annealed specimens were used in the present experiments and care was taken to avoid straining them by handling or by suddenly placing them in ambients at temperatures different from that of the room (Cooper and Wallace 1956). With potassium chloride 1 : 50 microsecond impulses were mainly used but measurements were also made with 1 : 5000 microsecond impulses at −195°c, 20°c and 220°c and with d.c. at 20°c. The average values of electric strength obtained with 1 : 5000 microsecond impulses were slightly less than the values obtained by testing comparable numbers of specimens with 1 : 50 microsecond impulses, but no significant difference existed in comparable maximum or minimum values. Consequently no distinction is made in subsequent sections of this paper between values of electric strength obtained with the different forms of voltage impulse, and the results of comparable measurements are lumped together. The difference in average values of electric strength mentioned above is consistent with the time lags in breakdown of a few microseconds duration observed by Cooper and Grossart (1956). With sodium chloride 1 : 5000 microsecond impulses were used throughout.

The results obtained with KCl and with NaCl are illustrated in figures 1 and 2 respectively, and on these diagrams are indicated the minimum values of electric stress under which a specimen may suffer permanent deformation of the crystal structure by the mechanical forces developed between the electric charges on the surfaces of the electrodes (Cooper and Wallace 1956). The numbers in brackets indicate the number of specimens tested. Specimens breaking down at values of electric strength above these lines contain considerable numbers of dislocations and other lattice imperfections which increase in density with the electric field. Since any anisotropy is expected to arise from the existence of a periodic crystal lattice structure weight must be given to the relative behaviour of those specimens with the most perfect structure. Reasons have been given (Cooper and Wallace 1956) for believing that the values of electric strength near to the minimum value correspond to these specimens. On this basis KCl is anisotropic at all temperatures below about 150°c, the (100) direction being the weakest and the (111) direction the

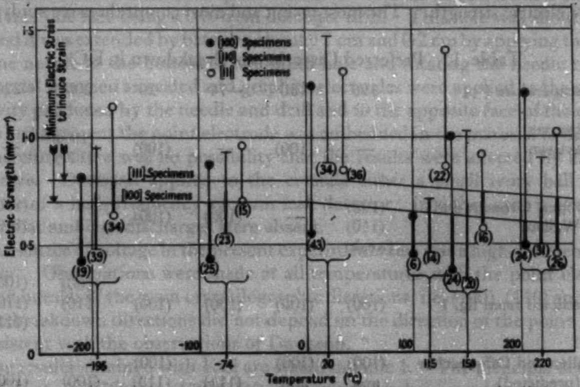


Figure 1. Electric strength of potassium chloride.

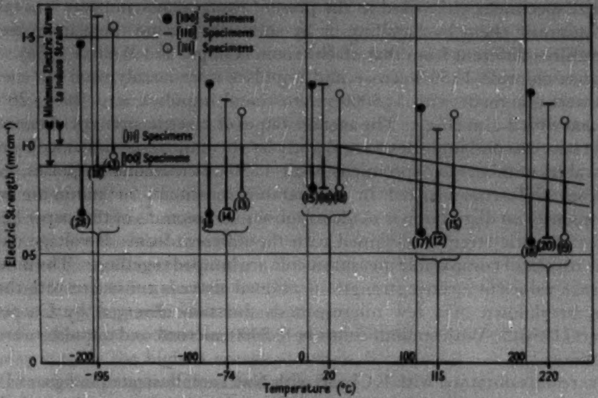


Figure 2. Electric strength of sodium chloride.

strongest. In view of the preceding discussion it is interesting to notice in figure 1 that evidence of anisotropy provided by comparing the maximum values of electric strength is perhaps less definite than is that obtained by comparing the minimum values. Except at −195°c there is very little evidence of anisotropy in NaCl, the small differences in minimum values at −74°c, 115°c and 220°c being hardly significant.

§ 3. PREFERRED CRYSTALLOGRAPHIC DIRECTIONS OF BREAKDOWN TRACKS

If the discharge tracks originating from a point electrode indicate the weakest directions in the crystal for electrical breakdown then agreement should exist between observed path directions and those predicted from measurements of intrinsic electric strength. The most recent and most comprehensive study of the

Table 1. Preferred Directions of Breakdown in KCl

Temperature (°c)	−195	−180	−74	20	200	220
Observer Davisson	—	(100)	—	(100)	(100) (110)	—
Cooper, Grossart and Wallace	(100) (110) (weak)	—	(100)	(100)		(100) (110)
Predicted from fig. 1	(100)	(100)	(100)	(100)	(100) (110)	(100) (110) (111)
Callen and Offenbacher (predicted)	(100) and random	(100) and random	(111)	(100) (110)	(100)	(100)

Table 2. Preferred Directions of Breakdown in NaCl

	Temperature (°c)						
Observer	−195	−180	−74	20	115	200	220
Davisson	—	(100)	—	(100) (111)	—	(100) (110)	
Cooper, Grossart and Wallace	$(100)_{10}$ $(110)_{1}$	—	$(110)_{4}$ $(100)_{1}$	$(110)_{4}$ $(100)_{1}$	$(110)_{15}$ $(100)_{5}$ $(111)_{1}$	—	$(110)_{4}$ $(100)_{1}$
Predicted from fig. 2	(100)	(100)	(100) (110)	(100) (110) (111)	(110) (100)	(100) (110) (111)	(100) (110) (111)
Callen and Offenbacher	Random or (100)	Random or (110)	Random and (100)	(110)	(100)	(100)	(100)

discharge tracks has been made by Davisson (1946) and in tables 1 and 2 his observations are compared with the predictions of our intrinsic strength measurements for KCl and NaCl respectively. Some differences exist, particularly for NaCl. According to Davisson the track directions depend both upon temperature and upon the impurities in the crystal. The temperatures used by Davisson were not the ones found most convenient for our intrinsic strength measurements nor is it certain that our specimens were chemically identical with his. It is also possible that internal strains in the crystal may influence the track directions because they do not develop homogeneously about a point when sufficient pressure is applied (Smakula and Klein 1951). Nothing is known about the state of strain of Davisson's specimens. It was therefore considered desirable to repeat the observations, using pieces of strain-free crystal obtained from the same parent crystals from which specimens were obtained for the electric strength measurements and at the same temperatures as these measurements.

The specimens consisted of plates about 0·75 cm thick and a blind hole was drilled into one face using a 0·076 cm diameter drill. This hole was about 0·2 cm deep and it was extended by between about 0·1 cm and 0·2 cm by applying the point of a fine needle to the bottom of the blind hole and rotating the needle rapidly. The crystal was then annealed and graphite electrodes were applied to the sides of the cavity produced by the needle and drill and to the opposite face of the crystal. In this arrangement the point electrode was embedded in the material to be tested and therefore there was no possibility that the results were affected by ambient discharges. Davisson applied to the crystals either a small ivory ball or the electrode of a high frequency vacuum leak detector; in either case it cannot be certain that ambient discharges were absent.

The source of voltage in the present experiments also was a high-frequency leak detector. Observations were made at all temperatures with the point electrode directed in each of the main crystallographic directions, i.e. (100), (110) and (111), but the breakdown directions did not depend on the direction of the point. This is consistent with the observations of Davisson.

The results obtained with KCl are stated in table 1. The principal discharge direction at −195°c was (100) but occasionally subsidiary (110) tracks were observed. No (111) tracks were observed with this material at −74°c and thus no support is given to the prediction of Callen and Offenbacher. The results obtained with NaCl are stated in table 2. Very infrequently (111) tracks were observed at 115°c, but at no other temperature were they ever observed, including 0°c which is within the range of temperatures (extending from about $-50°$c to about $+15°$c) characterized according to the theory of Callen and Offenbacher by preference for (111) discharges. Sometimes more than one track direction appeared at a given temperature in the same specimen and the subscripts in tables 1 and 2 indicate the relative frequency of occurrence of each particular track direction. Just as in the case of the intrinsic strength measurements, anisotropy is more definite in KCl than it is in NaCl.

In the majority of cases the discharge tracks lay in (100) crystallographic directions. Since mechanical failure in the alkali halide crystals generally occurs by cleavage in the (100) planes, it is perhaps important to give two reasons for believing that the observed discharge tracks resulted from the passage of an electrical discharge and not from mechanical failure of the crystal arising from the high mechanical pressure which must exist at the tip of an incipient discharge. In the first place, oscillograms show that the collapse of voltage across a specimen generally takes place in less than 10^{-8} second, the limit of resolution of the apparatus used. This time is compatible with that required for the production of a molten channel by the passage of an electron avalanche. It is incompatible with the time required for the production of a conducting path through a crack caused by interaction of shock waves, having a velocity of perhaps 2000 metres per second (Kolsky 1953) and originating at the tip of the incipient electrical discharge. Secondly, examination of the discharge tracks in a high-power ($\times 500$) microscope shows them to be approximately cylindrical channels from which the material has been removed and not the two-dimensional cleavages usually associated with mechanical failure in these materials.

§ 4. DISCUSSION

There is very good agreement shown in table 1 between the discharge track directions observed in KCl and those predicted from the intrinsic strength measurements. This depends to some extent, but not entirely, on accepting the low values of electric strength as the significant ones. This practice is unusual in this field of work. Low values of electric strength are generally regarded with suspicion because such values can easily occur if insufficient care is taken with the experimental techniques. For example, ambient discharges may cause premature breakdown; failure may occur from electric stress concentrations at small cracks within the specimens or at surface irregularities. It is reasonably certain that the results of the present experiments were free from effects of ambient discharges since, within the limits imposed by the wide spread, there was no dependence of electric strength on specimen thickness or upon the ambient medium. It is also reasonably certain that the specimens were free from macroscopic cracks or incipient cleavages. Such defects are easily detected by rotating the specimen in a beam of light, because strong reflection occurs from the defect at the appropriate angle and thus an easily detected scintillation is produced. It can, of course, still be maintained that other smaller surface irregularities cause field concentrations and hence low values of electric strength. This opinion is difficult to disprove by direct experiment, but measurements recently made by Smith (1956) suggest that the microscopic irregularities in the crystal surfaces do not influence electric strength.

In recent years some attention has been given to the possibility of space-charge accumulation giving rise to low values of electric strength. For example, some ionic migration may occur under the influence of the strong electric field and produce a space charge which significantly modifies the distribution of field within the crystal. Alternatively, field emission of electrons from the cathode may result in the same effect. The magnitude of both of these spurious effects is likely to differ specimen by specimen and, consequently, if they are present, appreciable spread in values of electric strength may result, but in this case the highest values of electric strength are likely to be the significant ones. The existence of these effects has been suggested on the evidence of an apparent influence of impulse duration on electric strength (Alger and von Hippel 1949, Inuishi and Suita 1953, 1954), but it is possible to propose an alternative explanation of the effects observed, at least below about 150°c, based on the appreciable time lags which exist in the breakdown of alkali halide crystals (Cooper and Grossart 1956) and upon possible statistical effects arising from the wide spread in possible values and the small numbers of specimens which were apparently used by the above authors. But unless ionic conductivity is strongly anisotropic, and there is no evidence of this, It is unlikely that the differences in electric strength shown in figure 1 can be accounted for in the above fashion.

Agreement between the observed discharge tracks in KCl and those expected from electric strength measurements is good up to 220°c, at which temperature the intrinsic strength measurements indicate that (111) tracks should be observed, but this was not the case. There appears to be a change in the mechanism of breakdown in this material at about 120°c. Below this temperature the electric strength increases slowly with rise in temperature in a manner consistent with a mechanism of breakdown dependent on interactions between electrons and lattice vibrations, but above this temperature the electric strength no longer continues to increase. The effect is most marked with the specimens of highest electric strength, i.e. the (111) specimens for which the fall in electric strength between 120°c and 150°c is precipitous. It is not believed that a critical temperature exists at about 120°c, above which breakdown is determined by electron–electron interactions in the manner proposed by Fröhlich (1939). The cause of the behaviour above 120°c is not known, but time lag measurements (Cooper and Grossart 1956) have provided evidence of the flow of very appreciable pre-breakdown current at 220°c and it is thought that breakdown at the higher temperature was influenced by thermal effects. The same situation exists in NaCl, but to a lesser extent. It is possible that the high internal impedance of the leak detector prevented the flow of appreciable pre-breakdown current form the point electrode and, consequently, only electronic breakdown phenomena were observed with this device.

Sodium chloride is much less anisotropic than is potassium chloride and there are some differences between the track directions expected from intrinsic strength measurements and those actually observed. The main differences lie in the prediction of (111) tracks at 20°c, 200°c and 220°c and in the appearance of (111) tracks at 115°c, although only (110) and (100) tracks were predicted. When a comparison of this kind is made, it must be realized that experimental errors make it difficult to be certain about the reality of apparent differences in electric strength of about ±10% or less, but small differences of this order may be decisive in determining the direction of the discharge tracks. Consequently, little significance can be attached to the above differences. The agreement is better at low temperatures, where the anisotropy in NaCl is more pronounced. The belief that alkali halide crystals are isotropic with respect to electric strength must be seriously questioned and, consequently, doubts must exist about the validity of a theory of breakdown based on the properties of low-energy conduction electrons, as proposed by von Hippel and developed by Callen (1949).

The theory due to Callen and Offenbacher (1953) represents, as far as we know, the first attempt to account quantitatively for the phenomenon of preferential discharge direction. Despite the apparent agreement of its predictions with the observations of Davisson it cannot be deemed satisfactory. Although our observations were made at slightly different temperatures than were Davisson's they are substantially in agreement with his, but the failure to observe any (111) tracks in KCl at −74°c and in NaCl at 0°c is more important than apparent agreement at temperatures where the (100) direction is preferred. On the basis of the scattering probability calculations we would expect the greatest electric strength to be observed, if breakdown is determined by the properties of fast conduction electrons, when the field is applied in the (111) direction and, on the same basis, the least values of electric strength when it is in the (100) direction. Our observations are not inconsistent with this.

ACKNOWLEDGMENTS

The experiments were carried out in the Electrical Engineering Department of Liverpool University, with the aid of maintenance grants from the Department of Scientific and Industrial Research and from Liverpool University, to whom we wish to offer our thanks.

REFERENCES

ALGER, R. S., and VON HIPPEL, A., 1949, *Phys. Rev.*, **76**, 127.

CALDERWOOD, J. H., COOPER, R., and WALLACE, A. A., 1953, *Proc. Instn Elect. Engrs*, **100**, Pt. IIA, 105.

CALLEN, H., 1949, *Phys. Rev.*, **76**, 1374.

CALLEN, H., and OFFENBACHER, F. L., 1953, *Phys. Rev.*, **90**, 401.

COOPER, R., and GROSSART, D. T., 1956, *Proc. Phys. Soc.* B, **69**, 1351.

COOPER, R., and WALLACE, A. A., 1956, *Proc. Phys. Soc.* B, **69**, 1287.

DAVISSON, J. W., 1946, *Phys. Rev.*, **70**, 685.

DAVISSON, J. W., and VON HIPPEL, A., 1940, *Phys. Rev.*, **57**, 156.

FRÖHLICH, H., 1939, *Rep. Progr. Phys.*, **6**, 41 (London : Physical Society); 1947, *Proc. Roy. Soc.* A, **63**, 81.

INUISHI, Y., and SUITA, T., 1953, *Tech. Rep. Osaka Univ.*, **3**, No. 80; 1954, *Ibid.*, **4**, No. 112.

KOLSKY, H., 1953, *Stress Waves in Solids* (Oxford : University Press), Chapter VIII.

SMAKULA, A., and KLEIN, H., 1951, *Phys. Rev.*, **84**, 1043.

SMITH, W., 1956, *Proc. Phys. Soc.* B, **69**, 848.

Electrical and Optical Properties of Type IIb Diamonds

By P. T. WEDEPOHL†

Department of Physics, University of Reading

Communicated by R. W. Ditchburn; MS. received 10th October 1956, and in final form 7th November 1956

Abstract. The d.c. Hall coefficient and resistivity of six semiconducting diamonds have been measured as a function of temperature from 200 to 800°K. All the diamonds were p-type semiconductors. From the measurements the effective mass of the charge carriers, the concentration of donor and acceptor centres, the activation energy of the acceptors (close to 0·34 ev) and the temperature dependence of the Hall mobility have been calculated. At high temperatures the mobility varies as $T^{-2\cdot 8}$.

The relative strengths of certain peaks in the infra-red absorption spectrum are the same in different specimens. It is concluded that all these peaks are due to the same centre. These peaks are not detected in type I and type IIa diamonds. There is a strong correlation between the strength of the strongest peak at 0·346 ev and the number of vacant acceptor levels at room temperature, from which it is concluded that the centre giving rise to the peaks is also responsible for conduction.

---

§ 1. INTRODUCTION

CUSTERS (1952) discovered that some diamonds which transmitted ultra-violet light (type II) were also semiconductors. He referred to these as type IIb diamonds. Since 1952 several workers have investigated the electrical and optical properties of type IIb diamonds (Leivo and Smoluchowski 1955, Custers 1955, Brophy 1955, Austin and Wolfe 1956, Dyer and Wedepohl 1956). In each case the measurements were carried out on a single specimen only.

This paper reports measurements of Hall coefficient, resistivity and infra-red absorption made on six type IIb diamonds, in order to investigate possible correlations of these properties. All specimens were p-type semiconductors, and had bulk resistivities lying between 50 and 1200 ohm cm at room temperature. All the diamonds had been cut into rectangular blocks having approximate dimensions 4 mm × 2 mm × 2 mm.

§ 2. HALL COEFFICIENT AND RESISTIVITY MEASUREMENTS

Hall coefficient and resistivity have been measured as a function of temperature from about 200 to 800°K. Standard d.c. potentiometric methods were used. Contact to the diamond was made by coating the appropriate faces or spots with a suspension of colloidal graphite in water. The external connections were made by spring-loaded copper probes or plates bearing on the graphite. The magnetic field used for measuring Hall coefficient could be determined with a probable error of 2%. The fields used were in the range 2700 to 3000 oersteds. High temperature measurements were carried out under vacuum to prevent oxidation of the diamond surface.

---

† On leave from Diamond Research Laboratory, Johannesburg

When measuring resistance as a function of temperature, the distance between probes was kept constant. These resistance values were multiplied by a factor ρ_0/r to convert them to resistivity values, where ρ_0 is the mean value of the room temperature resistivity, and r is the resistance between probes at room temperature.

To determine the mean value of the room temperature resistivity, the resistance was measured for several different probe separations. With three specimens the scatter about the mean value was greater than could be accounted for by experimental error (±5%). The extra scatter is probably due to some inhomogeneity in the distribution of centres responsible for conduction. Brophy (1955) also found evidence for inhomogeneity in the diamond investigated by him.

§ 3. Results

The Hall coefficient R was calculated from the measured Hall voltage, magnetic field and specimen current, and the hole concentration p was then found from

$$R=\frac{1}{p\mathbf{e}}\frac{\mu_{\mathrm{H}}}{\mu_{\mathrm{D}}} \qquad \ldots\ldots(1)$$

where $\mathbf{e}$ is the electronic charge, μ_{H} the Hall mobility and μ_{D} the drift mobility.

The usual correction for specimens having a finite length : width ratio was applied to the Hall voltages when calculating the Hall coefficients (Isenberg, Russell and Greene 1948).

From simple semiconductor theory (e.g. Shockley 1950), it can be shown that $\mu_{\mathrm{H}}/\mu_{\mathrm{D}}=3\pi/8$. This theory presumes that lattice scattering predominates, and that the crystal minimum energy surfaces are single spheres. The second assumption is probably incorrect, and the first is probably valid for the diamonds investigated only at temperatures above 400°K. For simplicity we have assumed that $\mu_{\mathrm{H}}/\mu_{\mathrm{D}}=3\pi/8$ throughout, since insufficient information is available to determine the true value of $\mu_{\mathrm{H}}/\mu_{\mathrm{D}}$ (which probably varies with temperature).

Figures 1 and 2 show hole concentration, Hall coefficient and resistivity as a function of temperature for one specimen (D105).

§ 4. Mobility

The resistivity ρ is given by

$$\frac{1}{\rho}=p\mathbf{e}\mu_{\mathrm{D}}. \qquad \ldots\ldots(2)$$

With equation (1) this gives

$$\mu_{\mathrm{H}}=R/\rho. \qquad \ldots\ldots(3)$$

Figure 3 shows Hall mobility as a function of temperature for five diamonds, the scale being logarithmic. The curves all show a similar temperature dependence at high temperatures, where lattice scattering probably predominates. In this region μ_{H} varies approximately as $T^{-2\cdot 8}$. These results are similar to those reported recently for the lattice mobility of electrons in silicon, and of holes in germanium and silicon, which give a temperature dependence of $T^{-2\cdot 3}$ to $T^{-2\cdot 7}$ (Prince 1953, Morin and Maita 1954, Ludwig and Watters 1956).

At lower temperatures the curves fall below the $T^{-2\cdot 8}$ line, probably due to the onset of impurity scattering.

Earlier theoretical calculations of the lattice mobility give a temperature dependence of $T^{-1\cdot 5}$ (e.g. Bardeen and Shockley 1950). These calculations are based on the assumption that the energy band edge lies at the centre of the Brillouin zone, and that there is a quadratic relation between energy and crystal momentum.

The absolute value of the mobility at any temperature depends directly on the resistivity (see equation (3)), which depends on the value of the resistivity measured at room temperature. This value was difficult to determine precisely, due to the inhomogeneity mentioned before. The uncertainty in the resistivity values could account for the variation in mobility from specimen to specimen at high temperatures, where it would be expected to be constant for a given temperature.

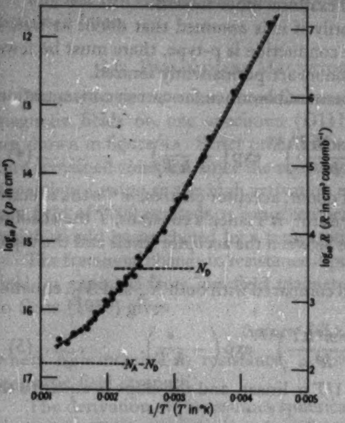


Figure 1. Hole concentration and Hall coefficient as a function of inverse absolute temperature for D105.

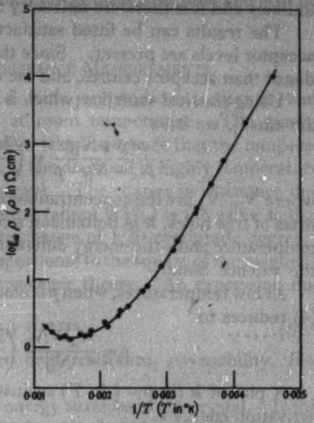


Figure 2. Resistivity as a function of inverse absolute temperature for D105.

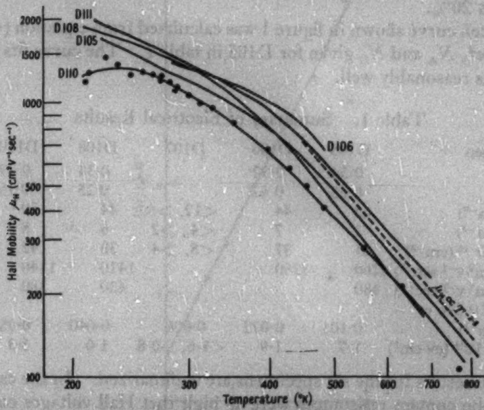


Figure 3. Hall mobility as a function of temperature for five diamonds. Experimental points are given for one specimen (D110) to indicate scatter.

§ 5. Analysis of Concentration Data

The simplest model for an extrinsic p-type semiconductor is one in which there is a single acceptor level above the valence band. In order to fit the experimental results to the equation derived on this model it is necessary to assume that the effective mass of the free holes is about 300 free electron masses. A similar result was found by Austin and Wolfe (1956). This effective mass is extremely unlikely and it is therefore necessary to examine other models.

The results can be fitted satisfactorily if it is assumed that donor as well as acceptor levels are present. Since the conduction is p-type, there must be fewer donor than acceptor centres, and the donors are permanently ionized.

Using classical statistics (which is permissible, since the carrier concentrations are small), we have

$$\frac{p(p+N_{\mathrm{D}})}{N_{\mathrm{A}}-N_{\mathrm{D}}-p}=\left(\frac{2\pi m^{*}kT}{h^{2}}\right)^{3/2}\exp\left(-\frac{\epsilon}{kT}\right) \qquad \ldots\ldots(4)$$

where N_{D}, N_{A} are the concentrations of donor, acceptor centres, m^* is the effective mass of free holes, k is Boltzmann's constant, h Planck's constant, T the absolute temperature and ϵ the energy difference between the acceptor levels and the top of the valence band.

At low temperatures, when p is small compared with both N_{A} and N_{D}, equation (4) reduces to

$$p=\frac{N_{\mathrm{A}}-N_{\mathrm{D}}}{N_{\mathrm{D}}}\left(\frac{2\pi m^{*}kT}{h^{2}}\right)^{3/2}\exp\left(-\frac{\epsilon}{kT}\right). \qquad \ldots\ldots(5)$$

A plot of $k(\ln p-\frac{3}{2}\ln T)$ against $1/T$ is linear, and the slope is equal to the activation energy ϵ.

With the value of ϵ found as above, N_{A}, N_{D} and m^* were calculated from the results over the complete temperature range to give the best fit to equation (4). The nature of the equation is such that although there is an uncertainty of as much as a factor two in the values for N_{D} and m^*, ϵ could be determined to within 5% and N_{A} to within 20%.

The broken curve shown in figure 1 was calculated from equation (4), with the values of ϵ, m^*, N_{A} and N_{D} given for D105 in table 1. The curve fits the experimental points reasonably well.

Table 1. Summary of Electrical Results

Specimen	D105	D106	D107	D108	D110	D111
ϵ (ev)	0·34	0·32		0·34	0·29	0·34
m^*/m_0	0·25	0·62		0·25	0·16	0·25
$N_{\mathrm{A}}\times10^{-15}$ (cm^{-3})	63	44	<12, >6	44	50	35
$N_{\mathrm{D}}\times10^{-15}$ (cm^{-3})	2	7	<4, >2	6	8	3
$(N_{\mathrm{A}}-N_{\mathrm{D}})\times10^{-15}$ (cm^{-3})	61	37	<8, >4	38	42	32
μ_{H} at 290°K (cm^2 v^{-1} sec^{-1})	1260	1380		1410	1140	1450
μ_{H} at 500°K (cm^2 v^{-1} sec^{-1})	480			430	380	560
I = Strength of 0·346 ev peak (ev cm^{-1})	0·101	0·072	0·006	0·040	0·056	0·039
$I/(N_{\mathrm{A}}-N_{\mathrm{D}})\times10^{18}$ (ev cm^2)	1·7	1·9	<1·6, >0·8	1·0	1·3	1·2

In table 1 the results for the six specimens are summarized. In the case of D107 the Hall probe contact resistances were so high that Hall voltages could not be measured. Approximate values for the carrier concentration were calculated by combining measured values of resistivity with average mobility values for the other diamonds. The same method was used for D106 above 450°K.

The values of the activation energies are close to 0·34 ev, with the exception of that of diamond D110. The ($\log p$, $1/T$) curve for this specimen is non-linear and less temperature dependent below 250°K, whereas the curves for the other specimens are all linear down to 200°K. Possibly there is an additional acceptor level in D110 very close to the valence band. This would falsify the value of the activation energy found in the usual way.

The last two rows in the table refer to the optical measurements, and will be discussed later.

§ 6. PRELIMINARY MAGNETORESISTANCE MEASUREMENTS

Magnetoresistance has been measured with perpendicular electric and magnetic fields on one specimen (D111) at room temperature. The results are shown in figure 4. Strict precautions were taken to ensure that the temperature remained constant, since the resistance of the diamond is highly temperature dependent (owing to the high activation energy). The change in resistance due to a field of 2000 oersteds at room temperature (about 0·3%) is the same as that which would be produced by a temperature change of 1/20 centigrade degree.

The fractional change in resistance is proportional to the square of the magnetic field, as expected from low field magnetoresistance theory. An expression due to Seitz (1950) gives

$$\Delta\rho/\rho_0 = 3{\cdot}8 \times 10^{-17}\mu^2 H^2 \qquad \ldots\ldots(6)$$

where $\Delta\rho$ = increase in resistance, ρ_0 = zero field resistance, μ = mobility, H = magnetic field strength.

The derivation of (6) assumes spherical energy surfaces in the crystal. If this assumption is correct, then μ given by equation (6) should be equal to the Hall mobility. As yet measurements have been made using only one direction of

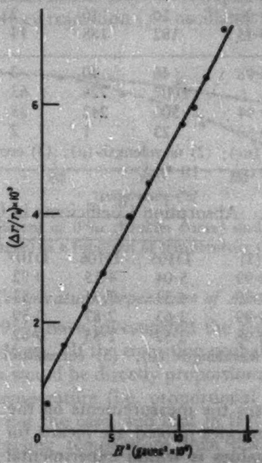


Figure 4. Magnetoresistive effect. The fractional increase in resistance is plotted against the square of the magnetic field (D111).

current and one of magnetic field. For this case μ (calculated from (6)) is 3800 cm^2 v^{-1} sec^{-1}. This is approximately three times the Hall mobility.

The plot shown in figure 4 is non-linear at low fields and apparently does not pass through the origin. This could be due to the possibility that the line joining the resistance probe contacts was not exactly parallel to the electric field, which would lead to a small Hall potential difference between the probes. The effect is being investigated.

§ 7. OPTICAL MEASUREMENTS

7.1. *Absorption Spectra*

The infra-red absorption spectra of the six diamonds have been investigated at room temperature, using a Grubb–Parsons double-beam infra-red spectrophotometer, between 1 and 10 μ.

The main features of the absorption spectra in the region investigated (see, for example, Clark, Ditchburn and Dyer 1956) are: (*a*) Broad absorption bands in the regions of about 3, 4 and 4·8 μ: these bands are also present in type IIa diamonds. (*b*) A very broad absorption band with short wavelength tail, extending from about 0·5 to 3 μ: this band is not detected in type IIa diamonds, and will be referred to as the 0·5–3 μ band. (*c*) Eight easily resolved absorption peaks, some of which are superimposed on the 0·5–3 μ band. These peaks are not detected in type IIa diamonds. The wavelengths and the corresponding energies of the peak maxima are given in table 2 (*a*).

Table 2 (*a*). Infra-red Peak Strengths

			(4)					
(1)	(2)	(3)	D105	D106	D107	D108	D110	D111
0·675	1·84	0·87	33	30	4	19	34	18
0·537	2·31	0·90	16	16	2	5	11	4
0·514	2·41	0·95	182	188	14	95	133	94
0·372	3·33							
0·363	3·42	0·98	46	40	3	16	22	17
0·346	3·58		1007	724	63	404	556	389
0·307	4·04	0·94	300	242	35	125	259	146
0·295	4·20	0·68	23	1	2	2	1	

(1) Mean energy of peak (ev); (2) wavelength (μ); (3) correlation coefficient; (4) integrated absorption of peak (ev cm$^{-1} \times 10^{-4}$).

Table 2 (*b*). Absorption Coefficients in 0·5–3 μ Band

			(4)					
(1)	(2)	(3)	D105	D106	D107	D108	D110	D111
0·6	2·07	0·99	5·04	4·15	0·52	2·12	3·06	2·36
0·7	1·77	0·99	4·34	3·54	0·53	1·89	2·93	2·13
0·8	1·55	0·99	3·63	2·87	0·59	1·72	2·55	1·89
1·0	1·24	0·98	2·52	1·85	0·62	1·33	1·89	1·47

(1) Energy (ev); (2) wavelength (μ); (3) correlation coefficient; (4) absorption coefficient (cm^{-1}).

Table 2 (*a*) summarizes the measurements on the infra-red peaks of the six diamonds. The peak energies quoted are means for the six specimens. The variation in individual values is within experimental error. When measuring integrated absorption of the peaks, allowance was made for specimen thickness.

It is not possible to make quantitative measurements on the strength of the 0·372 ev peak, since it is too weak in all specimens investigated.

Linear correlation coefficients between the strengths of the 0·346 ev peaks and those of the other peaks have been calculated to investigate possible relationships between them (see table 2(*a*)). It is concluded that the peaks at 0·675, 0·537, 0·514, 0·363, 0·346 and 0·307 ev arise from transitions in the same absorbing centre. The peak at 0·295 ev does not appear to be associated with this group.

Correlation coefficients between the strength of the 0·346 ev peak, and absorption coefficients at four wavelengths in the 0·5–3 μ band, have also been calculated (see table 2(*b*)). The correlation is very good, from which it is concluded that the 0·5–3 μ band is also due to the centre responsible for the six peaks mentioned above.

7.2. *Correlation of* 0·346 ev *Peak Strength with Number of Vacant Acceptor Levels*

The values obtained for the electrical activation energy ($\simeq$0·34 ev) are close to the energy corresponding to the strongest infra-red peak (0·346 ev). This suggests that the absorption of 0·346 ev is due to the excitation of electrons from discrete levels just above the valence band to the vacant acceptor levels. In this case the integrated absorption at 0·346 ev in different specimens should be proportional to the density of vacant acceptor levels.

At room temperature a negligible fraction of the acceptor centres is ionized, so that the density of vacant acceptors is nearly equal to $N_A - N_D$. The last two rows in table 1 give the integrated absorption of the 0·346 ev peak, and the ratio of this to the value of $N_A - N_D$. These ratios are constant within experimental error.

It is concluded that the centre giving rise to the conduction is also responsible for the absorption peaks at 0·346, 0·675, 0·537, 0·514, 0·363 and 0·307 ev, and for the 0·5–3·0 μ absorption band.

The ratios given in the last line of table 1 correspond to an *f*-value of between 0·01 and 0·02 for the 0·346 ev transition (f = oscillator strength).

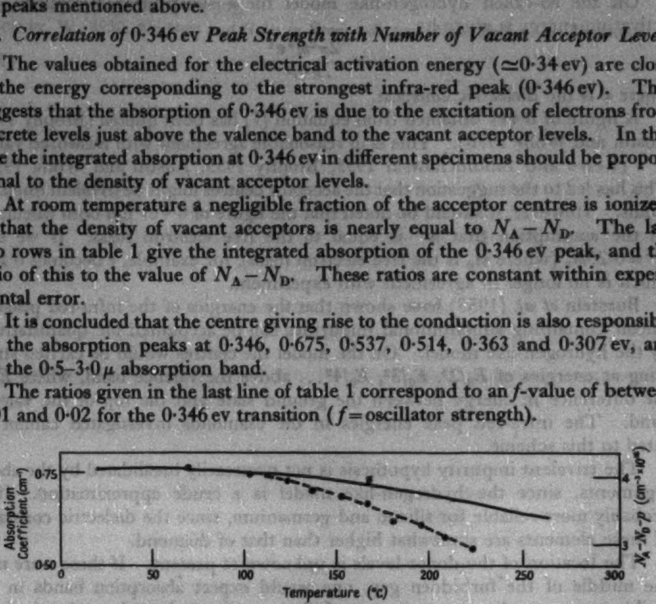


Figure 5. Absorption coefficient at 0·9μ (broken curve) and density of vacant acceptors (full curve) as a function of temperature (D110).

7.3. *Temperature Dependence of Absorption*

Austin and Wolfe (1956) have investigated the temperature dependence of the absorption line at 0·346 ev. If the transition probability were constant, then the integrated absorption would be directly proportional to the number of vacant acceptor levels at any temperature (i.e. proportional to $N_A - N_D - p$). Their results show that this is not the case, and that in fact the absorption decreases more sharply with increasing temperature than does the number of vacant acceptor levels.

We have measured the absorption coefficient at 0·9 μ in the 0·5–3 μ band as a function of temperature for one specimen (D110), from room temperature to about 200°c. The results are plotted in figure 5. It can be seen that in this case, too, the absorption coefficient is more temperature dependent than the density of vacant acceptor levels.

It is concluded that, although the centre responsible for conduction is also responsible for the 0·5–3 μ absorption band, the temperature dependence of this absorption cannot be accounted for quantitatively by a simple theory based on constant transition probability.

§ 8. DISCUSSION

On the so-called hydrogen-like model for a substitutional impurity, the activation energy is given by

$$\epsilon = \frac{2\pi^2 m^* e^4}{\kappa^2 h^2} \qquad \ldots\ldots(7)$$

where κ is the dielectric constant.

A figure of 0·4 ev obtained from equation (7) has been quoted (Brophy 1956, Austin and Wolfe 1956). This is in reasonable agreement with measured values of ϵ (Leivo and Smoluchowski 1955, Brophy 1955, Austin and Wolfe 1956). This has led to the suggestion that the acceptor centres might be trivalent impurity atoms. However, it should be noted that the figure of 0·4 ev has been calculated on the assumption that m^* is equal to the free electron mass. If we take $m^* = 0{\cdot}25 m_e$ (where m_e is the free electron mass), for example, ϵ becomes 0·1 ev, which is no longer in agreement with experiment.

Burstein *et al.* (1953) have shown that the energies of the infra-red peaks in silicon containing substitutional impurities can be accounted for quantitatively by the hydrogen-like model. On the model the centres would be excited states lying at energies of $E_0/2^2$, $E_0/3^2$, $E_0/4^2\ldots$ above the valence band, where E_0 is the difference in energy between the ground state and the top of the valence band. The infra-red peak energies in the diamonds investigated cannot be fitted to this scheme.

The trivalent impurity hypothesis is not necessarily invalidated by the above arguments, since the hydrogen-like model is a crude approximation. It is probably more reliable for silicon and germanium, since the dielectric constants of these elements are somewhat higher than that of diamond.

The location of the donor levels is unknown at present. If these were near the middle of the forbidden gap, one would expect absorption bands in the visible, corresponding to excitation of electrons from the valence band to the empty donor levels. The only absorption which has been detected in the visible is the short-wavelength tail of the 0·5–3·0 μ band. As shown above the strength of this tail correlates with the concentration of unoccupied acceptor centres.

ACKNOWLEDGMENTS

The author wishes to thank Messrs. Industrial Distributors (1946) Ltd. for supporting this research, Dr. J. F. H. Custers, Director of Research, Diamond Research Laboratory, Johannesburg, for his help and co-operation and for permission to publish this paper, Dr. Goulden of the National Institute for Research in Dairying, Shinfield, who carried out the infra-red measurements, and Professor R. W. Ditchburn, Dr. C. D. Clark, Dr. H. B. Dyer and Dr. E. W. J. Mitchell for encouragement and helpful advice.

REFERENCES

AUSTIN, I. G., and WOLFE, R., 1956, *Proc. Phys. Soc.* B, **69**, 329.
BARDEEN, J., and SHOCKLEY, W., 1950, *Phys. Rev.*, **80**, 72.
BROPHY, J. J., 1955, *Phys. Rev.*, **99**, 1336; 1956, *Physica*, **22**, 156.
BURSTEIN, E., BELL, E. E., DAVISSON, J. W., and LAX, M., 1953, *J. Phys. Chem.*, 57, 849.
CLARK, C. D., DITCHBURN, R. W., and DYER, H. B., 1956, *Proc. Roy. Soc.* A, **234**, 363.
CUSTERS, J. F. H., 1952, *Physica*, **18**, 489; 1955, *Nature, Lond.*, **176**, 173.
DYER, H. B., and WEDEPOHL, P. T., 1956, *Proc. Phys. Soc.* B, **69**, 410.
ISENBERG, J., RUSSELL, B. R., and GREENE, R. F., 1948, *Rev. Sci. Instrum.*, **19**, 685.
LEIVO, W. J., and SMOLUCHOWSKI, R., 1955, *Bull. Amer. Phys. Soc.*, **30**, No. 2, 9.
LUDWIG, G. W., and WATTERS, R. L., 1956, *Phys. Rev.*, **101**, 1699.
MORIN, F. J., and MAITA, J. P., 1954, *Phys. Rev.*, **96**, 28.
PRINCE, M., 1953, *Phys. Rev.*, **92**, 681, *Ibid.*, **93**, 1204.
SEITZ, F., 1950, *Phys. Rev.*, **79**, 372.
SHOCKLEY, W., 1950, *Electrons and Holes in Semiconductors* (New York: van Nostrand).

On the Complex Permeability of Iron–Nickel Alloys at High Frequencies

By J. C. ANDERSON † and B. DONOVAN ‡

† Department of Physics, Northern Polytechnic, London.

‡ Department of Physics, Bedford College, London.

MS. received 2nd November 1956, *and in final form 4th December* 1956

Abstract. The real and imaginary parts of the complex permeability have been measured for a series of iron–nickel alloys in the frequency range 300–400 Mc/s. The frequency of the internal resonance observable in this region has been investigated as a function of composition and is found to pass through a minimum in the neighbourhood of 70% Ni. The value of the principal anisotropy coefficient is also known to be very small in this region and a qualitative comparison provides additional support for the view that this resonance is attributable to the equivalent magnetic field associated with the anisotropy energy.

---

§ 1. INTRODUCTION

ALTHOUGH the phenomenon of ferromagnetic resonance in an external magnetic field has been extensively investigated in recent years, the possibility has been appreciated, since the early paper of Landau and Lifshitz (1935), that an internal resonance should also be observable in ferromagnetic materials. In this effect the elementary domains, in the absence of an external field, may be regarded as moving in the internal field of the crystal, arising from the anisotropy energy. For crystal fields of the order of 100 oersteds, resonance would be anticipated at frequencies around 300 Mc/s. An anomaly in the permeability of an iron–nickel wire was observed in this region by Wieberdink (1948, 1950) and was interpreted as an internal resonance by Kronig (1948). An experimental investigation of the corresponding effect in ferrites was carried out by Birks (1950).

Special interest attaches to the problem in the case of the iron–nickel alloys in view of their many characteristic magnetic properties. In particular, the anisotropy coefficient, upon which the internal field depends, is known to vary markedly with composition and the measurements described in the present paper were undertaken to determine the resonant frequency for a series of specimens, containing 36% to 100% nickel, and to discover the extent of its dependence upon composition. Extensive measurements on single crystals (see, for example, Bozorth 1952, 1953) have shown that the principal anisotropy coefficient changes sign in the neighbourhood of 70% nickel, and it might, therefore, be expected that, even for partially ordered specimens, the resonant frequency would pass through a minimum in this region. This expectation is, in fact, borne out by the results presented here and the resonant frequency exhibits a relatively shallow but significant minimum in the region of low anisotropy energy.

§ 2. EXPERIMENTAL DETAILS

2.1. *Method*

The apparatus used to measure the complex permeability was similar to that described by Wieberdink (1950, 1951), involving a coaxial transmission line in which the specimen under test constitutes the inner conductor. The line was approximately one metre long and the outer conductor was a silver plated copper tube of internal diameter 3·82 cm. The specimens were either wires of approximate diameter 0·3 mm or thin strips of width 0·6–1·2 mm and thickness 0·05–0·3 mm. A second coaxial line with both conductors of copper was used as a wavemeter to obtain the free-space wavelength. All measurements were carried out at room temperature.

The formulation of the complex permeability and its relation to experimental measurement has been discussed by Kittel (1946). In the present technique one requires the internal impedance per unit length of the inner conductor, together with its d.c. resistance, and also the external self inductance per unit length. If the permeability is written as $\mu = \mu_1 - i\mu_2$ it is possible to obtain expressions for μ_1 and μ_2 in terms of the wavelength on the line, the propagation constant and the free-space wavelength. For a cylindrical conductor the relevant theory has been given by Wieberdink (1950, equations (21)–(29)). In the case of the strip specimens the internal impedance is readily obtained, if the skin depth is assumed to be small compared with the dimensions of the specimen (which is valid in all the measurements described here); the external self inductance was found by the method of conformal transformation.

2.2. *Specimens*

Thirteen specimens were investigated, with compositions varying from 36% to 100% nickel, and the details are given in the accompanying table. Several of the alloys were standard commercial products; specimens h–m were supplied to special order by the Mond Nickel Company.

To ensure uniformity the heat treatment was standardized as far as possible, viz. annealing at 1000°c in dry hydrogen for one hour, subsequently cooling to room temperature in 10 hours. The exceptions were specimens e, f and g, the H.C.R. (grain oriented) strip being cold rolled after annealing and Nilo 65 being supplied as a hard-drawn wire.

Permalloy C was specified as containing 3% molybdenum and the nickel

Details of Specimens

Designation	% Ni	Form	Description	Source	Resonant frequency (Mc/s)
a	36	strip	Permalloy D	S.T.C.	336
b	45	„	Permalloy B	S.T.C.	328
c	80·5	„	Permalloy C	S.T.C.	323
d	50	wire	Nilo 50	M.N.	327
e	65	„	Nilo 65	M.N.	318
f	50	strip	H.C.R.	T.C.C.	325
g	84·5	„	Mumetal	T.C.C.	321
h	70	„	Special order	M.N. rolled and heat treated by T.C.C.	324
j	75	„			331
k	80	„			324
l	85	„			332
m	90	„			342
n	100	wire	Pure Nickel	M.N.	378

M.N. = Mond Nickel Co., S.T.C. = Standard Telephones and Cables, T.C.C. = Telegraph Construction Co.

percentage entered in the table represents the ratio Ni/(Ni + Fe). The same proportion was used in the case of Mumetal, for which the exact constitution (and heat treatment) was not divulged and the Ni–Fe ratio only was supplied.

§ 3. Discussion of Results

Using the method of Wieberdink (1950) the real and imaginary parts of the permeability have been derived for each of the specimens listed in the table and a typical pair of curves is shown in figure 1, in which μ_1 and μ_2 are plotted as functions of the free-space wavelength λ_0 for Permalloy D. Superimposed on the well known decrease of permeability with frequency, a characteristic irregularity is encountered in this frequency range which, as was originally remarked by Wieberdink, is strongly reminiscent of the phenomenon of anomalous dispersion. This behaviour may be interpreted as an internal resonance occurring in the anisotropy field of the specimen and it should be pointed out that the resonance peak is exhibited by μ_2, the imaginary part of the permeability, which is derived from the real part of the internal impedance of the specimen.

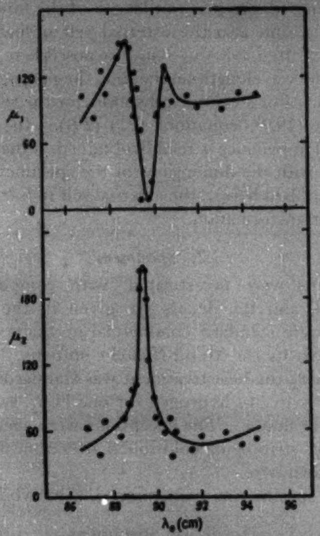


Figure 1. Variation of real and imaginary parts of complex permeability with free-space wavelength in neighbourhood of internal resonance. (Specimen a: 36% Ni.)

In view of this possible interpretation it is of particular interest to examine the variation of the resonant frequency with composition. The accuracy with which the frequency may be determined depends to some extent on the sharpness of resonance but in general may be reckoned to be about 2 per cent. In figure 2 the resonant frequency is shown as a function of the percentage of nickel for the range of specimens investigated. The letter beneath the curve represents the specimens listed in the table and the additional point, labelled w, represents Wieberdink's result for a wire specimen containing 60% nickel. Although the overall variation in frequency is not large there is nevertheless a significant general trend, with a minimum value at about 65% nickel and a small subsidiary maximum in the neighbourhood of 75% nickel. It will be seen that the result for Mumetal lies appreciably below the main curve.

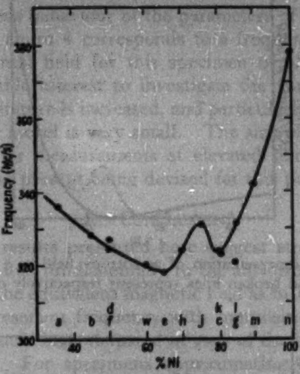


Figure 2. Variation of resonant frequency with composition. The letters indicate the various specimens as listed in the table; Wieberdink's result for 60% Ni is identified by the letter w.

Although it is not possible to provide a quantitative explanation of these results a useful qualitative insight may be obtained by considering the variation of the anisotropy field. The fundamental quantity is the energy associated with the crystal anisotropy and for cubic crystals the predominant term in the energy density may be expressed as

$$E = K_1(\alpha_1{}^2\alpha_2{}^2 + \alpha_2{}^2\alpha_3{}^2 + \alpha_3{}^2\alpha_1{}^2),$$

where the α's are the direction cosines of the magnetization vector with respect to the cubic axes and K_1 is the principal anisotropy coefficient. The influence of this anisotropy on magnetic resonance has been discussed by Kittel (1948) for single crystals with specific orientations and the effect may be regarded as equivalent to a magnetic field, which for cubic crystals is of the order of $2K_1/I_s$ where I_s is the saturation magnetization.

Measurements on iron–nickel single crystals at room temperature (Bozorth 1952, Chap. XII; 1953) show that K_1 changes sign at a composition which varies from 63% to 75% nickel, depending on the rate of cooling. Using the data given by Bozorth, $|K_1|/I_s$ may be evaluated as a function of composition and the result is shown in figure 3. The full line applies to a quenched specimen, while the broken line is obtained with a very slow rate of cooling (2 deg h^{-1}) which leads to a considerable degree of ordering and is responsible for the subsidiary maximum in the neighbourhood of FeNi$_3$. It will be seen that the general form of the curves in figure 3 is strikingly similar to that of the experimental results shown in figure 2, and the minimum resonant frequency does, in fact, occur in the region of zero anisotropy. It is somewhat surprising that the curve in figure 2 should exhibit a small hump of the same type and in the same position as that in figure 3. With the high rate of cooling ($\sim$100 deg h^{-1}) used in the present specimens it seems highly unlikely that any significant irregularity of this kind would be observed in K_1 (cf. Bozorth 1953, figure 5) or that any appreciable degree of ordering would be obtained. Nevertheless, this hump was found to be perfectly reproducible and the fact that it occurred with a range of specimens (h–m inclusive) with identical treatment encourages the view that it is caused by a genuine variation of the anisotropy energy.

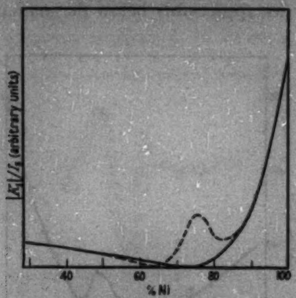


Figure 3. Approximate representation of anisotropy field as a function of composition. The continuous and broken lines represent respectively quenched and very slowly cooled specimens.

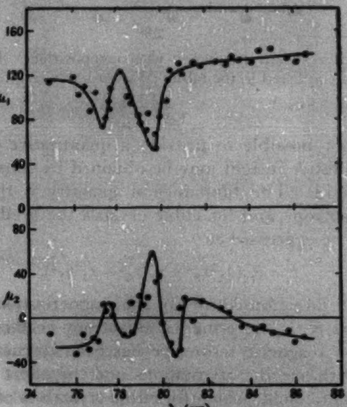


Figure 4. Variations of real and imaginary parts of complex permeability with free-space wavelength for pure nickel specimen.

Particular interest attaches to the resonance observed in pure nickel which, so far as the authors are aware, has not previously been reported. The behaviour of the complex permeability is shown in figure 4 and is considerably more complicated than the straightforward resonance peaks obtained with the other specimens. The pattern suggests a splitting of the main frequency but the cause of such fine structure seems at present rather obscure. The negative values of μ_2 have no special physical significance; they appear naturally as a consequence of the measurements (cf. Wieberdink 1950, equation (28)) and do not imply any anomalous behaviour of the parameters of the coaxial line.

The main peak in figure 4 corresponds to a frequency of 378 Mc/s which indicates a mean internal field for this specimen of about 140 oersteds. It would be of considerable interest to investigate the variation of the resonant frequency as the temperature is increased, and particularly in the neighbourhood of 100°c where K_1 for nickel is very small. The size of the present apparatus renders it unsuitable for measurements at elevated temperatures and a more convenient method is at present being devised for this purpose.

§ 4. Conclusion

The experimental results presented here suggest strongly that the internal resonance observed is due fundamentally to the anisotropy energy and may be visualized in terms of the equivalent magnetic field as in Kronig's interpretation. The variation of the resonant frequency with composition, although not large, is of the form which would be anticipated in view of the known behaviour of the anisotropy coefficient. For specimens approximating more closely to single crystals it might be expected that the minimum frequency would be very much lower; in the case of a very extended frequency range the present method would no longer be suitable.

It is hoped to extend these measurements by using a modified technique to investigate single crystals and hence, from a more precise knowledge of the anisotropy field, to obtain information of quantitative significance on the internal resonance.

References

Birks, J. B., 1950, *Proc. Phys. Soc.* B, **63,** 65.

Bozorth, R. M., 1952, *Ferromagnetism* (New York: van Nostrand); 1953, *Rev. Mod. Phys.*, **25,** 42.

Kittel, C., 1946, *Phys. Rev.*, **70,** 281; 1948, *Ibid.*, **73,** 155.

Kronig, R., 1948, *Nature, Lond.*, **162,** 528.

Landau, L., and Lifshitz, E., 1935, *Phys. Z. Sowjet*, **8,** 153.

Wieberdink, A., 1948, *Nature, Lond.*, **162,** 527: 1950, *Appl. Sci. Res.* B, **1,** 439; 1951, *Physica*, **17,** 333.

The Influence of Boundary Conditions on the Onsager–Ising Partition Function for the Plane Square Lattice

By H. N. V. TEMPERLEY

Atomic Weapons Research Establishment, Aldermaston, Berks.

MS. received 19*th September* 1956 *and in revised form* 19*th October* 1956

Abstract. A formal proof is given that, in the critical region, the predictions of the model are the same whether the atoms are considered to be arranged on a torus, on a cylinder or on a plane. The latter arrangements involve ‘raw’ edges, and are hard to handle by matrix methods, but can be dealt with by a slight modification of Kac and Ward’s treatment. The conclusion is not trivial, because critical assemblies are known to be abnormally sensitive to perturbations, and it is in principle possible that, since correlation distances are large, an alteration in the boundary constraints might affect the entire assembly.

---

§ 1. INTRODUCTION

Up to the present, all investigations of the Onsager–Ising model have assumed, implicitly or explicitly, that the atomic magnets are arranged on a torus, an assumption equivalent to that of periodic boundary conditions. For a review of the methods see Newell and Montroll (1953). In the matrix method of approach the matrix V describing the effect of adding another row of N atoms to the lattice, is defined in such a way that the first and Nth atoms are considered to be nearest neighbours. One then proceeds to calculate $\mathrm{Tr}(V^M)$. The diagonal elements of V^M describe configurations in which the first and last rows are precisely equivalent to one another, while the non-diagonal elements describe other configurations. As Onsager pointed out in his original paper (1944), the true partition function for atoms arranged on a *cylinder* would be given by the sum over *all* the elements of V^M, which is a difficult quantity to compute because, unlike the trace, it is *not* invariant under the transformations TVT^{-1}. It *does* seem to have been established, Kaufman and Onsager (1949), that, for physical purposes, the trace can be replaced by the Mth power of the top eigenvalue, but it has never been proved that, if the first and last rows are independent of one another, the true partition function can be effectively replaced by the trace of V^M.

It is this point that we shall investigate in this paper, using, for the purpose, a slight modification of the determinant of Kac and Ward (1952) and it will be found that, near the critical temperature, the partition function is *not* significantly modified if periodic boundary conditions are replaced by the assumption that two or four of the edges are ‘raw’. This conclusion is not trivial, and by no means follows from the argument that the Ising model has proved suitable in practice for representing actual physical assemblies known to be unaffected by boundary conditions. Indeed, various other pieces of information about co-operative assemblies would be consistent with the idea that the properties in the critical region *were* altered, for example: (*a*) It has been pointed out by many writers (see, for example Tisza

1951), that an assembly near its Curie point is very sensitive to small disturbances, for example fluctuations are becoming large and correlation effects are extending over larger and larger distances. The latter effect is shown by the Ising model itself (Kaufman and Onsager 1949). Schneider and Atack (1951) have shown experimentally that the effect of gravity can become significant in the critical region of a gas, although it is trivial elsewhere.

(*b*) It has become evident that, even in the very simple Weiss model of a phase transition, the passage to the limit of a large assembly requires great care (Temperley 1954, 1956).

It is therefore *possible* that the effect of a constraint, such as the requirement that the first and last rows of a lattice must have identical configurations, may not always be localized, but may extend through the entire assembly. The argument given by some writers, that the boundary effect on the partition function becomes relatively small as $N, M \to \infty$, may not always be valid.

§ 2. THE KAC AND WARD DETERMINANT

We refer to Kac and Ward's original paper (1952) for a proof that the calculation of the partition function for the plane square lattice is equivalent to the evaluation of a certain determinant, which can be carried out by multiplying together all the eigenvalues of the corresponding matrix. The $4NM \times 4NM$ determinant has two rows or columns for each of the $2NM$ links, each link corresponding to one of the possible ways of joining up two nearest neighbours in the lattice, and the evaluation of the partition function, known to be equivalent to the enumeration of the number of different patterns made up of closed domains of a given total perimeter, is shown to be equivalent to evaluation of the determinant. The fact that a domain can be encircled in either the clockwise or the counter-clockwise sense merely means that the determinant is really describing two independent $N \times M$ assemblies simultaneously. (This holds rigorously only if we may neglect 'domains' whose perimeters cross the boundary, but we are in any case going to eliminate these in our modified treatment.)

The eigenvalue problem is equivalent to the solution of a set of simultaneous equations, of which we give some specimens below. The determinant to be evaluated has units all down the main diagonal, each such factor corresponding to a link that is not used. The non-diagonal elements correspond in row to the link that has just been used, in column to one of the three possible ones that may follow. x is the variable whose power counts the total number of links in each pattern, while $\alpha^4 = -1$, the point of inserting the α's being fully explained by Kac and Ward (1952):

$$\left.\begin{aligned}(\lambda-1)R_{n,m} &= xR_{n+1,m} + x\alpha U_{n+1,m} + x\alpha^{-1}D_{n+1,m}\\ (\lambda-1)U_{n,m} &= x\alpha^{-1}R_{n,m+1} + xU_{n,m+1} + x\alpha L_{n,m+1}\\ (\lambda-1)L_{n,m} &= x\alpha^{-1}U_{n-1,m} + xL_{n-1,m} + x\alpha D_{n-1,m}\\ (\lambda-1)D_{n,m} &= x\alpha R_{n,m-1} + x\alpha^{-1}L_{n,m-1} + xD_{n,m-1}.\end{aligned}\right\} \quad \ldots\ldots(1)$$

Now, if the links are arranged in a completely cyclic fashion, equations (1) contain the following:

$$\left.\begin{aligned}(\lambda-1)R_{N,m} &= xR_{1,m} + x\alpha U_{1,m} + x\alpha^{-1}D_{1,m},\\ (\lambda-1)U_{n,M} &= x\alpha^{-1}R_{n,1} + xU_{n,1} + x\alpha L_{n,1},\end{aligned}\right\} \quad \ldots\ldots(2)$$

and a similar set involving L and D on the left-hand side, and it is possible to satisfy equations (1) simply by putting

$$R_{n,m}=R\delta^n\epsilon^m,\quad U_{n,m}=U\delta^n\epsilon^m,\ \text{etc., with}\ \delta^N=\epsilon^M=1, \qquad \ldots\ldots(3)$$

all $4NM$ eigenvalues being thus obtained, as each possible choice of δ and ϵ gives rise to a quartic equation for λ, the constant term of which gives the product of the four eigenvalues, which products are themselves multiplied together to evaluate the determinant.

§ 3. EFFECT OF MODIFYING THE BOUNDARY CONDITIONS

We now examine the effect of forbidding the use of all links like $R_{N,m}$, $U_{n,M}$, $L_{1,m}$, $D_{n,1}$, which corresponds to deleting all terms on the right-hand sides of equations like (2). We are retaining links like $U_{N,m}$, $D_{N,m}$ as these correspond to upward or downward links on the right-hand boundary, and we also retain links that lead *inwards* from the boundary, while we are not using links that project *outwards* from the boundary at right angles. We know that each possible choice of links corresponds to an arrangement of atoms in the dual lattice rather than the actual lattice, and a little consideration shows how the cyclic and non-cyclic models differ. According to the cyclic model, the dual lattice contains $N+2$ columns but column 0 is constrained to have the same configuration as column N, while column 1 is constrained to be identical with $N+1$, the interactions between atoms above one another in columns 0 and $N+1$ being neglected. According to the non-cyclic model, interactions between atoms in 0 and 1 and between N and $N+1$ are considered but there is no other constraint on the configurations of these two outside columns, and interactions *within* them are neglected.

The new problem is to evaluate the determinant whose completely cyclic character has been destroyed by replacing by zero the terms corresponding to all those 'links' that would be used by domains crossing the boundary. We look for eigenfunctions of the corresponding new matrix, that, instead of satisfying equations (2), have now the property that $R_{N,m}$, $U_{n,M}$ etc. all vanish. Equations (1) are equivalent to a set of *difference* equations and, guided by the analogy with ordinary boundary-value problems, we attempt to solve our present problem by superposing solutions, of the *form* (3), to the corresponding problem involving completely cyclic conditions. This turns out to be possible, provided that we make suitable modifications in δ, ϵ. (In some elementary problems the 'cyclic' solutions can be superposed as they stand, but cases in which the parameters they contain have to be modified before the solutions can be so used are also common.) We examine this point by trying to satisfy equations (1), together with the new boundary conditions replacing equations (2), by solutions of the form

$$R_{n,m}=A\delta^n\epsilon^m+B\delta^{-n}\epsilon^m+C\delta^n\epsilon^{-m}+D\delta^{-n}\epsilon^{-m}, \qquad \ldots\ldots(4)$$

(and similarly for $L_{n,m}$, etc.), where δ, ϵ are no longer roots of unity, but are quantities that have to be determined. These two changes are based on standard methods of solving boundary-value problems. Our conditions require that $R_{N,m}$ should vanish, for all m and for any ϵ that leads to a solution, which in general requires that $A\delta^N+B\delta^{-N}$ and $C\delta^N+D\delta^{-N}$ must vanish separately. Similarly, the vanishing of $L_{1,m}$ requires the vanishing of $E\delta+F\delta^{-1}$, where E, F are related to A, B in such a way that the 'main' equations (1), i.e. those which remain in (1) after removal of the 'boundary' equations (2), are satisfied by $R_{n,m}=A\delta^n\epsilon^m$, $L_{n,m}=E\delta^n\epsilon^m$, etc., separately. Substituting these elementary solutions in the main equations, we obtain the eigenvalue equation

$$[(1-\lambda)^2+x^2]^2+x(1-\lambda)[(1-\lambda)^2-x^2](\delta+\delta^{-1}+\epsilon+\epsilon^{-1})=0 \quad \ldots\ldots(6)$$

and the following relations between the coefficients:

$$\frac{A}{E}=-\frac{F}{B}=\alpha^2\frac{(1-\lambda)^2\delta+(1-\lambda)x(1+\delta\epsilon^{-1})-x^2\epsilon^{-1}}{(1-\lambda)^2\delta^{-1}+(1-\lambda)x(1+\delta^{-1}\epsilon^{-1})-x^2\epsilon^{-1}}$$

$$=\alpha^{-2}\frac{(1-\lambda)^2\delta+(1-\lambda)x(1+\delta\epsilon)-x^2\epsilon}{(1-\lambda)^2\delta^{-1}+(1-\lambda)x(1+\delta^{-1}\epsilon)-x^2\epsilon}. \quad \ldots\ldots(7)$$

Then, since $A\delta^N+B\delta^{-N}$ and $E\delta+F\delta^{-1}$ must vanish to satisfy the boundary conditions, we get the following equation for δ:

$$\pm\delta^{N-1}=iA/E. \quad \ldots\ldots(8)$$

It is possible to eliminate the ratio $(1-\lambda)/x$ between (6), (7) and (8) to get an explicit form for the relation between δ, ϵ and M, N but it does not seem possible to get this into a simple form. At first sight, since the possible values of δ seem to be linked with those of ϵ, whereas in the torus case these two variables are independent of one another, it would seem that the distribution of possible values of δ and ϵ in the two problems is quite different, with a corresponding effect on the partition function. Indeed, *some* such effect is to be expected. We have only to make N and M small for such a modification to become necessary.

We can, however, show that, in the important region where δ and ϵ are both near ± 1, the distribution of values is of the same form in both problems. Consider first the case of atoms arranged on a *cylinder*, so that there are only *two* raw edges. In this case the variable ϵ is still a root of unity, since the boundary conditions are still periodic in the corresponding direction. From equations (7) and (8) we have, taking $\alpha^2=i$ and the positive sign in (8):

$$\frac{\delta^N-\delta}{\delta^{N-1}-\delta^2}=-\frac{(1-\lambda)^2+(1-\lambda)x\epsilon^{-1}}{(1-\lambda)x-x^2\epsilon^{-1}}, \quad \ldots\ldots(9)$$

$$\frac{\delta^N+\delta}{\delta^{N-1}+\delta^2}=\frac{(1-\lambda)^2+(1-\lambda)x\epsilon}{(1-\lambda)x-x^2\epsilon}. \quad \ldots\ldots(10)$$

Now if $\epsilon=\pm 1$, (9) and (10) imply $\delta^{2N-1}-\delta^3=0$, which means that the distribution of the δ's is the same as for an assembly of effective size of the order of $2N$, with periodic boundary conditions. If ϵ is near ± 1, it can be written $\pm 1+[O(1)]/M$, and this only implies an unimportant correction in the δ's, which were formerly $\pm 1+[O(1)]/N$ and now have an additional term $[O(1)]/MN$. Since the important contributions to the partition function near the Curie temperature come, as we shall see below, from the region where δ and ϵ are nearly equal to ± 1, which quantities have now been shown to differ only insignificantly from roots of unity in this region, we conclude that the changes in the partition function are also unimportant, at least in the region of temperature near the Curie point.

We can go over without difficulty from the case of two 'raw' edges to that of four 'raw' edges. We now get a pair of equations for ϵ as well as δ and it is again possible to see that, in the important region ϵ, $\delta\simeq\pm 1$, the distribution of values differs insignificantly from that for the periodic case. For we can solve the equations by an iteration process. We have seen that, in the important region, if ϵ is a root of unity, δ only differs from one by a quantity of the order $1/MN$, and this remains true if such a correction is applied to ϵ also, further iterations only adding terms to δ and ϵ involving still higher inverse powers of M and N.

Once any particular set of δ's and ϵ's are determined, we can proceed exactly in the manner indicated by Kac and Ward (1952) for the evaluation of the partition function. For each pair, the product of the four corresponding values of λ is given by the constant term in (6), that is the well-known expression

$$(1+x^2)^2+x(1-x^2)(\delta+\delta^{-1}+\epsilon+\epsilon^{-1}), \qquad \ldots\ldots(11)$$

so that the complete log (partition function) is obtained by taking the logarithm of (11) and summing over all δ, ϵ. Near the critical region, (11) becomes very small for δ, ϵ near -1, in which circumstances $\delta+\delta^{-1}$ differs from -2 in the periodic case by a term of the order n^2/N^2 (n an integer), while $\epsilon+\epsilon^{-1}$ contains a similar term of the order m^2/M^2, and the form of the anomaly follows correctly if we sum the logarithm of (11) over m, n rejecting the terms involving higher inverse powers of M and N, which vanish when the assembly becomes large. When we go over the 'raw edges' case, the modifications in δ and ϵ mean, first that M, N are replaced effectively by $2M$, $2N$ (a modification that is familiar enough in boundary-value problems), which simply changes the partition function by a constant factor, while the corrections in δ, ϵ involving $1/MN$ introduce terms into (11) which are, at worst, of the order $1/M^2N$, $1/MN^2$, so are again negligible in comparison with the main terms.

We recall that the anomaly found by Onsager (1944) arises in the following way. The log (partition function) is obtained from the logarithm of expression (11) by summing over all possible δ, ϵ. The function $g(x)=x(1-x^2)/(1+x^2)^2$ is equal to $\frac{1}{4}$ at the critical value $(\sqrt{2}-1)$ of x and its first derivative vanishes there. The anomaly arises because expression (11) can vanish for $\delta=\epsilon=-1$, but does not become evident until we differentiate the log (partition function) *twice* with respect to x. In other words, the specific heat is, to a factor, given by

$$\text{Non-singular terms} \qquad +\sum\frac{g''(x)(\delta+\delta^{-1}+\epsilon+\epsilon^{-1})}{1+g(x)(\delta+\delta^{-1}+\epsilon+\epsilon^{-1})} \qquad \ldots\ldots(12)$$

and if δ, ϵ are equal to $-\exp(2m\pi i/M)$, $-\exp(2n\pi i/N)$ respectively, we find that the singular term in (12) is proportional to

$$\sum_{m,n}\left[\frac{m^2}{M^2}+\frac{n^2}{N^2}+\ldots\right]^{-1}$$

which, by comparison with the corresponding integral, is found to be proportional to $-MN\log(1/M^2+1/N^2)$, that is, the specific heat per atom increases logarithmically with the size of the assembly. The main contribution to this comes from the region where m and n are small, and would not be affected by changing the upper limits of the summation, nor, in the limit of large M and N, would the conclusion be changed if δ, ϵ were replaced by quantities differing from them only by quantities of the order $(MN)^{-1}$. Away from the critical region, the denominator of (12) never vanishes, and all values of δ and ϵ now contribute effectively to the answer.

We have thus succeeded in proving formally that the behaviour of the model in the critical region is unaffected, since this depends solely on the distribution of values of δ and ϵ near -1. As we recede from the critical region, the contributions for larger values of m and n become important, calling for a knowledge of the new values of δ, ϵ over an ever wider range. Since we have not determined the complete spectrum of values, the proof that the partition function is unaffected by the boundary conditions in the limiting case of a large assembly is mathematically incomplete. On physical grounds we can argue with some confidence that, since the form of the specific heat anomaly, which experience with approximate methods has shown to be a feature sensitive to small changes, is unaffected, the other physical consequences of the model are also likely to be preserved.

ACKNOWLEDGMENT.

I wish to thank the Director, Atomic Weapons Research Establishment, for permission to publish this paper.

REFERENCES

KAC, M., and WARD, J. C., 1952, *Phys. Rev.*, **88**, 1332.
KAUFMAN, B., and ONSAGER, L., 1949, *Phys. Rev.*, **76**, 1232, 1244.
NEWELL, G. F., and MONTROLL, E. W., 1953, *Rev. Mod. Phys.*, **25**, 353.
ONSAGER, L., 1944, *Phys. Rev.*, **65**, 117.
SCHNEIDER, W. G., and ATACK, D., 1951, *J. Phys. Chem.*, **55**, 532.
TEMPERLEY, H. N. V., 1954, *Proc. Phys. Soc.* A, **67**, 233; 1956, *Changes of State* (London : Cleaver-Hume Press), chapters 3 and 4.
TISZA, L., 1951, *Phase Transformations in Solids* (New York : Wiley), chapter 1.

The Variation of Friction with Velocity

By R. T. SPURR and T. P. NEWCOMB
Ferodo Ltd., Chapel-en-le-Frith, Stockport, Cheshire

Communicated by R. C. Parker; *MS. received 8th October* 1956

Abstract. It is shown that the friction of a bituminous material decreases with increasing sliding velocity, and that this decrease is due to the creep behaviour of the material.

§ 1. Introduction

Several kinds of friction velocity dependence have been described. Beare and Bowden (1935) and Bowden and Hughes (1939) found friction to be independent of velocity over a wide of range of speeds. According to Dokos (1946) and Forrester (1946) friction decreases with increasing velocity. ' Stick-slip ' motion has been thought to be due to such a decrease (Sampson *et al.* 1943). Maney (1952) examined the variation of friction with velocity by applying various tangential loads to the specimen and measuring its velocity if and when the latter became constant. Under these circumstances the friction appears to increase with increasing velocity.

It was observed that the friction of the bitumen 'Mexphalte' was markedly dependent upon the sliding velocity and this dependence has been examined in detail. Other aspects of the friction of Mexphalte have been examined previously (Spurr and Newcomb 1957). Mexphalte incidentally is a black material which has a Vickers hardness number of the order of 0·1 kg mm^{-2} at room temperature becoming markedly visco-elastic at higher temperature and is liquid at about 45°c.

§ 2. Experimental Details

The friction of small Mexphalte blocks (about 1 in. square) was measured on a friction and wear machine described elsewhere (Hughes and Spurr 1955). The speed of sliding could be varied over two ranges, viz. 5 to 25 cm sec^{-1} and 75 to 450 cm sec^{-1}. The specimens were 'run in' until the surfaces became smooth and polished before any measurements were made.

§ 3. Results and Discussion

The dynamic coefficient of friction μ_D was measured over the experimental range of speed using a normal load of 200 g. It was found that μ_D decreased with increasing speed. Typical results are shown in figure 1. When these results were plotted on a log–log scale (figure 2) it was found that up to speeds of about 100 cm sec^{-1} μ_D is given by the equation $\mu_D = K_1 v^{-0\cdot19}$ where v is the sliding speed. If it is assumed that friction is due to adhesive forces acting across the areas where the surfaces are in contact, and that the life t of the contact areas or junctions is inversely proportional to the sliding speed, then $\mu_D = K_2 t^{0\cdot19}$.

The static coefficient of friction of the Mexphalte μ_s depends upon the time t the surfaces are in contact before the measurement is made (Spurr and Newcomb 1957). To determine how μ_s varied with t the disc of the friction machine was clamped stationary and the specimen loaded (200 g load) against it for periods of time varying between 1 and 10 000 seconds and the friction measured. It was found that μ_s could be represented by the equation $\mu_s = K_s t^{0\cdot 10}$.

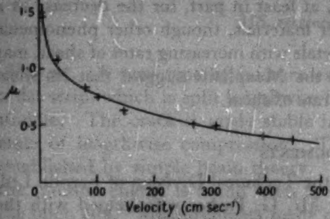


Figure 1. Variation of μ of Mexphalte with sliding velocity.

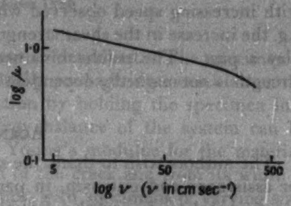


Figure 2. Logarithmic plot of curve of figure 1.

Thus, it appears that in both the static and dynamic cases, the friction of the Mexphalte is proportional to the same power of the life time of the contact areas. The time dependence of the static friction is due to creep increasing the real areas of contact, and, therefore, it is suggested that the friction velocity variation is also due to creep.

Further measurements were made which support this theory. The creep rate of the Mexphalte is very temperature dependent and consequently the effect of temperature on the friction velocity variation was examined. The disc of the friction machine was heated by an electrical resistance strip to a temperature of 30°c and the variation of the dynamic friction with velocity and of the static friction with time of application of the normal load measured. Again it was found that both the μ_s and the μ_D results at the lower speeds gave straight lines when plotted against t and v respectively on a log–log scale and that the numerical values of the slopes of the two lines was approximately the same (see table).

The increased creep rate of the Mexphalte at higher temperatures accounts for the curvature of the log μ_D–log v line at speeds above about 100 cm sec^{-1} for at such speeds the frictional heating of the Mexphalte surfaces becomes important. Thermocouples were embedded in the surface of a specimen, the temperature reached at various speeds was measured, tangents were drawn to the (log μ_D, log v) curve at speeds at which the thermocouples registered temperatures of 30 and 35°c and the slopes of these tangents determined. The values of the slopes are given in the table, together with the index n of the static friction equation $\mu_s = Kt^n$ with the plate at 30 and 35°c.

Specimen temp. (°c)	20°	30°	35°
Static index	0·19 (200)	0·30 (34)	0·40 (40)
Dynamic index Plate 20°c	0·19 (45)	0·30 (45)	0·50 (45)
Dynamic index Plate 30°c		0·32 (40)	0·45 (40)

The figures in brackets indicate the number of measurements on which the preceding figure is based.

Finally, static and dynamic measurements were made with loads of 550 g at room temperature. Altering the load changed the time dependence of the friction slightly, the static friction becoming proportional to $t^{0\cdot 21}$, the dynamic friction to $v^{-0\cdot 21}$. There was much more scatter in the experimental results at higher temperatures and loads because of damage to the specimen surface.

If we assume that the life of a contact area is inversely proportional to the sliding speed then the μ-velocity dependence of the Mexphalte is due entirely to creep. It seems likely that creep would account, at least in part, for the decrease in μ with increasing speed observed with other materials, though other phenomena, e.g. the increase in the shear strength of metals with increasing rates of shear, may play a part. The results obtained with the Mexphalte suggest that its shear strength is not markedly dependent on the rate of shear.

ACKNOWLEDGMENTS

The authors wish to thank the staff of the Technical Division of Ferodo Ltd. for assistance and criticism, in particular Mr. G. Parker who helped with the experimental work, and the directors of Ferodo Ltd. for permission to publish this paper.

REFERENCES

BEARE, B., and BOWDEN, F. P., 1935, *Phil. Trans. Roy. Soc.*, **234**, 329.
BOWDEN, F. P., and HUGHES, T. P., 1939, *Proc. Roy. Soc.* A, **172**, 263.
DOKOS, S. J., 1946, *J. Appl. Mech.*, **13**, A, 148.
FORRESTER, P. G., 1946, *Proc. Roy. Soc.* A, **187**, 439.
HUGHES, G., and SPURR, R. T., 1955, *Proc. Phys. Soc.* B, **68**, 106.
MANEY, C. A., 1952, *Amer. J. Phys.*, **20**, 203.
SAMPSON, J. B., *et al.*, 1943, *J. Appl. Phys.*, **14**, 689.
SPURR, R. T., and NEWCOMB, T. P., 1957, *Proc. Phys. Soc.* B, **70**, 98.

Experiments on Controlled Fractures

By J. J. BENBOW and F. C. ROESLER
Imperial Chemical Industries Ltd., Akers Research Laboratories, Welwyn, Herts.

MS. received 1*st October* 1956

Abstract. A new type of fracture experiment is described. The specimen is a flat strip, which is split lengthwise by gradually propagating a crack down the middle. The crack is made stable in direction by holding the specimen in a state of lengthwise compression. The energy balance of the system can be established by simple beam theory. Given Young's modulus for the material, the fracture energy is found from the length of the crack and from its width at the open end. Measurements made on plastics, with specimens of varying size, show that the crack propagation is governed by a constant specific fracture energy, which is invariant with respect to scale transformations of the stresses. For a sample of polymethylmethacrylate the specific fracture energy was $4{\cdot}9 \times 10^5$ dyn cm^{-1}. For a sample of polystyrene it was 25×10^5 dyn cm^{-1} for slow cracks, but dropped to 3×10^5 dyn cm^{-1} for fast-running cracks.

---

§ 1. INTRODUCTION

KNOWLEDGE of fracture processes is mainly circumstantial; it is limited in most cases to knowledge of the conditions which prevent fracture. These conditions are usually defined by the tensile strength, the yield strength, and other critical stress values. Such information may serve to assess what loads a body of given dimensions will safely bear, but it tells little about fracture as a physical process. Excessive loads produce, in most mechanical systems, catastrophic failure; this tends to involve dynamic phenomena, which are often too complicated for systematic observation and analysis. Closer investigation of the fracture process requires a system in which a fracture grows under well-defined, controlled conditions.

Several such systems have been described, but they are restricted in application. The splitting method of Obreimoff (1930) is based on the laminated structure of mica. The appliances used for rubber by Rivlin and Thomas (1953) were devised for highly extensible material. Conical indentation cracks (Roesler 1956) can only be made in very brittle material.

The apparatus described here can produce controlled fractures in the tougher type of solid represented by many engineering materials. It uses a wedge and a system of constraints, as described in detail later. The specimens are flat strips of material, which are split lengthwise gradually. The energy balance of the splitting process results in the relation

$$T = 3E\delta^2 b^3/(64s^4),$$

which is derived as equation (11) below. T is the specific fracture energy, E is Young's modulus, b is the width of the specimen, δ is the width of the crack at the end opened by the wedge, and s is the length of the crack. The equation is only valid for s/b much greater than unity, but, as is shown later, this causes no difficulty. In effect, provided E is measured separately, T may be determined from measurements of s, δ and b. By changing the size of the specimen the scale of the stress boundary conditions can be varied independently of the energy balance. The usefulness of the energy balance theory of crack growth can thus be tested.

This theory was originally proposed for glass (Griffith 1920, 1924), and the energy spent per unit area of newly formed crack surface was identified with the surface tension. For fractures in other, less brittle materials, the energy balance description will only be useful if the specific fracture energy, which is here much larger than the surface tension, can still be considered a constant of the material. The difference between total specific fracture energy and surface tension is spent irreversibly in a multitude of dissipative processes, such as plastic flow. Theoretically, this complicated interaction of elastic and dissipative processes at the crack edge is an unexplored problem. Thus theory gives, as yet, no clear answer to the question whether the specific fracture energy is a constant. Rivlin and Thomas argue that in certain cases such constancy should be expected. Empirically it is well established that the assumption of a constant specific fracture energy often gives, at least, a good approximation. Much relevant evidence is due to G. R. Irwin, J. A. Kies and W. T. George. Several authors have suggested that the energy balance description might even be useful for fast-running cracks.

The unexpected success of the energy balance theory, in cases for which it was not intended, warrants further investigation. There is also a certain practical utility in the determination of specific fracture energies for various materials.

§ 2. DEVELOPMENT OF THE APPARATUS

The design of mechanical systems for producing stable cracks is easier if displacements rather than stresses are the imposed boundary conditions. These displacements need to be controlled and measured accurately; therefore, they should be relatively large. This is an argument against systems in which uniform displacements are imposed on extended regions of the boundaries.

A wedge is a simple instrument for opening a crack gradually, and the displacements enforced by it are easy to measure. There is, however, a difficulty. Laminated material, such as the mica in Obreimoff's experiment, or grained material, such as wood, will be split cleanly by a wedge. But in isotropic material the crack made by a simple wedge is directionally unstable: it curves and runs to one side of the specimen. The effect can be demonstrated by attempting to split a strip of Perspex with hammer and chisel. A remedy had to be found.

Some partial success was obtained by restricting the motion on both sides of the crack to parallel, outward movement. The wedge was not applied to the material directly, but to a pair of clamps fastened to the edge of the specimen: the clamps were interlocked with guides so as to inhibit relative rotation. These sliding clamps produced straight cracks in square pieces of sheet, but with narrower specimens, in the form of strips, they were not effective. Nevertheless, the sliding motion of the clamps had given partial stabilization and it was therefore retained in the final apparatus.

Eventually it was found that the crack can be kept straight even in a narrow strip by subjecting the specimen, throughout the experiment, to a preset lengthwise compression. As the strip is free to expand sideways, the compression does not induce stress in that plane of symmetry of the system in which the crack is to run. Therefore, leaving aside higher-order effects which will be discussed later, the lengthwise compression does not interfere with the regular propagation of the crack. Obviously, however, the compression inhibits crack growth normal to the plane of symmetry, and so it must, to some extent, stabilize the crack direction. The strength of the effect depends on how the compression is applied. For maximum effect, the compressive forces transmitted through the halves of the split strip should be balanced by the apparatus and their sum should remain constant, independently of how the crack develops. With this ideal system the crack should be positively pulled back to the middle, even if at any stage a deviation does occur, for instance through outside action on the system.

The actual apparatus does not conform exactly to this ideal system, which would have required a more complicated design, incompatible with the use of sliding clamps. With the arrangement used, the balance of compression forces is not statically determinate: thus the degree of stability obtainable is hard to predict. In practice, however, the apparatus worked well, and the crack never ran to the side of the specimen while the compression was applied. In two out of three specimens the crack travelled within the central tenth of the total width; this was the criterion for accepting measurements, derived from an error estimate given later. In retrospect, the idea of preventing crack growth in certain directions by compression seems obvious, by analogy with the pre-stressing of concrete and with the toughening of glass.

The apparatus is shown in figure 1 (Plate). The specimen shown is the smallest size used, 1 inch wide and 6 inches long. All specimens were cut from sheets of commercial plastics $\frac{1}{4}$ inch thick; their widths varied from 1 inch to 4 inches. Specimens wider than 1 inch were 12 inches long; these required support against buckling. A certain minimum amount of compression is necessary for effective crack stabilization, and for the 12-inch specimens this is slightly above the buckling limit. The long specimens were, therefore, confined between guides, so that they could not leave the plane of the apparatus. The crack was started from a saw-cut; to avoid sudden cracking this cut was finished in the shape of a swallow-tail by making sloping incisions with a jeweller's saw, so that the cut is extended along the upper and the lower face of the specimen. A travelling microscope was used to measure the crack width at the clamps, to 1/100 millimetre, and the crack length was measured with a ruler, to the nearest millimetre. The wedge is driven by a screw and is thereby self-locking; it gives smooth control of the crack propagation, so that the spacing of readings is arbitrary. On the specimen shown in the photograph fifteen measurements were made.

§ 3. DERIVATION OF THE ENERGY BALANCE

Figure 2 shows the system to be analysed and explains the notation. The force P and the moment M are not measured by the apparatus; they occur in the derivations but not in the final formula for the specific surface energy. The crack width is exaggerated in the drawing; for the analysis it may be assumed that $\delta \ll b$, so that all deformations can be taken as small. The compression force Q will, for the moment, be left out of account; its influence on the energy balance will be considered in the next section.

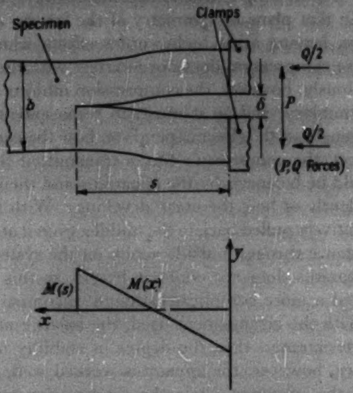


Figure 2. The system of the two beams and the moment distribution.

An inference concerning the form of the Griffith balance can be drawn from considerations of similarity. Any fixed value of s/b defines a manifold of geometrically similar systems. It follows, by an argument analogous to that given in an earlier paper (Roesler 1956), that the energy balance must be expressible in the form

$$\frac{T}{E} = \frac{\delta^2}{s}\,\phi\left(\frac{b}{s}\right) \qquad \ldots\ldots(1)$$

The argument assumes that the specific fracture energy T and Young's modulus E are constants of the material. The validity of equation (1) can be tested by plotting experimental values of $f(\delta^2/s)$ against $g(b/s)$, where f, g are any convenient functions. The plot should show coincidence of the results for specimens of different size.

For an absolute determination of T it is necessary to know the function $\phi(b/s)$ for at least one value of b/s, so that the plot of reduced variables can be calibrated. Now for s very much greater than b the two halves of the split strip can be regarded as long beams, which may be treated by the simple theory of flexure as used in engineering calculations. The result is the asymptotic form of $\phi(b/s)$, and since this is quickly approached in the plots it gives an adequate calibration.

According to the simple theory of flexure, the strain energy of a straight, uniform beam is

$$U = \frac{1}{2B}\int_0^s M^2(x)\,dx, \qquad \ldots\ldots(2)$$

where B is the stiffness given by $B = EI$, I is the moment of inertia of the cross section, $M(x)$ the bending moment, and x the distance along the beam. The boundary conditions of the present case require

$$M(x) = P(x - s/2), \qquad \ldots\ldots(3)$$

which is the distribution shown in figure 2. Substitution into (2) and integration results in

$$U(P,s)=P^2s^3/24B. \qquad \ldots\ldots(4)$$

Since δ is the measured quantity, P must be eliminated. By Castigliano's theorem the displacement of the beam end is

$$\delta/2=\partial U(P,s)/\partial P=Ps^3/12B, \qquad \ldots\ldots(5)$$

so that, solving for P,

$$P(\delta,s)=6B\delta/s^3. \qquad \ldots\ldots(6)$$

Substitution back into (4) then gives

$$U(\delta,s)=3\delta^2B/2s^3. \qquad \ldots\ldots(7)$$

This is the strain energy, in terms of the displacement, for one of the two beams. The fracture energy for one side of the crack is T times the area,

$$V(s)=Tns. \qquad \ldots\ldots(8)$$

The Griffith energy balance for the process of crack growth is

$$-\partial U(\delta,s)/\partial s=dV/ds, \qquad \ldots\ldots(9)$$

and with (7) and (8) it takes the form

$$Tn=9\delta^2B/2s^4. \qquad \ldots\ldots(10)$$

The moment of inertia of a rectangular cross section of height m and width n is given by $I=nm^3/12$; in the present case, with $m=b/2$, $I=nb^3/96$. The Griffith balance (10) can therefore be written, more explicitly,

$$T/E=3\delta^2b^3/64s^4. \qquad \ldots\ldots(11)$$

As would be expected, the thickness n of the specimen does not enter into this final equation for T. By comparison of (11) with (1) the previously unknown function $\phi(b/s)$ can now be identified: asymptotically $\phi=(3/64)(b/s)^3$.

To find the change of the force P during the splitting experiment the quantity $\delta(s)$, as defined by (10) or by (11), is put into (6). The result is

$$P(s)=(1/s)(8TnB)^{1/2}. \qquad \ldots\ldots(12)$$

Force and bending moment are related by (3), and for $x=s$, where the crack ends, the moment is given by $M(s)=Ps/2$. From (12) it then follows that, while the crack propagates, the bending moment $M(s)$ has a constant value

$$M(s)=(2TnB)^{1/2}=n(TEb^3/48)^{1/2}. \qquad \ldots\ldots(13)$$

Consequently, the stresses induced by $M(s)$ are constant. It follows that the repeated determination of T from successive values of δ and s for the same crack is no test of the energy criterion of fracture. It is the width of the specimen which must be varied to determine whether T is independent of variations in the stress field.

For specimens of varying b the stress at any characteristic point (say the point ahead of the crack edge at a distance $b/10$) will vary as $b^{-1/2}$, under the assumption that T is a constant. This can be shown from general similarity relations, as used to obtain equation (1). The same result can be found from equation (13) by using the expression for the maximum stress in the beam according to the simple theory, $\sigma=bM(s)/4I$. With equation (13) this gives

$$\sigma=(12ET/b)^{1/2}. \qquad \ldots\ldots(14)$$

This σ is a characteristic stress for the purpose of scaling relations. σ is not, however, the true stress at any specifiable point; in particular, it is not the stress at the crack edge. This assertion is perhaps obvious, but the two distinct reasons for it are worth formulating. First, the simple theory of flexure does not account for the stress concentration at the junction of the two beams. Secondly, the true stress at the crack edge cannot be found from elastic theory alone, even if the description of the elastic field surrounding the crack is available.

§ 4. INFLUENCE OF THE COMPRESSION, AND ERROR ESTIMATES

If the crack width were infinitely small the compression would not influence the energy balance. In the experiments δ/s had values up to 0·05; this is still small, yet large enough to require investigation of the error.

If a beam is loaded by transverse and by longitudinal forces then the theory must take into account both ordinary bending and buckling. So long as the longitudinal load Q is small compared with the critical load for buckling Q_c, the system can be treated by a perturbation method. This implies the use of a power expansion in Q/Q_c.

As stated earlier, the force Q employed in the experiments obeyed

$$Q \lesssim Q_c^* = 4\pi^2 B^*/L^2 = 4\pi^2 n^3 bE/12L^2, \qquad \ldots\ldots(15)$$

where Q_c^* is the lowest buckling load of the system. This pertains to the mode in which the specimen leaves the plane of the apparatus. L is the length of the specimen. Now for the buckling of the halves of the split strip in the plane of the apparatus the critical load is

$$Q_c = 2(4\pi^2 B/s^2) = 8\pi^2 nb^3 E/96s^2. \qquad \ldots\ldots(16)$$

It is this mode which interferes with the ordinary bending considered in the preceding section. Therefore, from (15) and (16),

$$Q/Q_c \lesssim 4(s/L)^2(n/b)^2. \qquad \ldots\ldots(17)$$

For the typical value $s = L/2$ the first two factors cancel. The last factor takes the value 1/144 for 3-inch strips of thickness $\frac{1}{4}$ inch. Thus Q/Q_c is truly a small parameter under the conditions of the experiment.

The leading term in the perturbation is obtained by assuming that the shape of the bent beam is determined by the transverse load alone. Bending diminishes the projection of the beam length on the x axis (figure 2); consequently, the external force Q puts work into the system. As $\cos x = 1 - x^2/2$ for small x, the change of the projected length is

$$\Delta = \int_0^s \left(\frac{dy}{dx}\right)^2 dx. \qquad \ldots\ldots(18)$$

The slope dy/dx is found by integrating the equation of the elastic line $d^2y/dx^2 = M(x)/B$, with $M(x)$ given by (3). Evaluation of (18) results in

$$\Delta = \frac{1}{240}\frac{P^2 s^5}{B^2} = \frac{3}{20}\frac{\delta^2}{s}. \qquad \ldots\ldots(19)$$

For one of the beams the total work of Q is $Q\Delta/2$. From (19) together with (4) or (7), this work is related to the strain energy U, as found earlier, by

$$\frac{Q\Delta}{2U} = \frac{1}{20}\frac{Qs^2}{B}. \qquad \ldots\ldots(20)$$

This relation shows the importance of the parameter Q/Q_0. The rate of work done by Q during crack growth, per unit area of one side of the crack, is given by

$$w = \frac{1}{n}\,\frac{Q}{2}\,\frac{d\Delta}{ds} = -\frac{3Q\delta^2}{40ns^2}. \qquad \ldots\ldots(21)$$

The growth condition (11) can now be used to eliminate δ. For further interpretation Q must also be eliminated. It seems plausible that the value of Q which will make the crack stable should be expressible in terms of the stresses. Using the characteristic stress (14), the Q to be imposed can be written

$$Q = nb\eta\sigma, \qquad \ldots\ldots(22)$$

where η is a numerical constant. Empirically it was found that $\eta \simeq 0{\cdot}2$ ensures stability. Thus by using (11) and (22) the error can be written

$$\frac{w}{T} = -\frac{16}{10}\left(\frac{s}{b}\right)^2 \eta \left(\frac{12T}{Eb}\right)^{1/2}. \qquad \ldots\ldots(23)$$

The right-hand side of (23) can be evaluated by inserting for s, b, E and η empirical data, and for T the value determined by (11). This gave $w/T = -0{\cdot}046$ for the narrowest specimen ($b = 1$ inch) and the longest crack ($s/b = 4$) of the measurements on polymethylmethacrylate. In most cases, however, the error term w/T was smaller, of the order $-0{\cdot}01$. Since w is negative the supply of strain energy must be increased to propagate the crack; consequently, the error in T as determined from (11) is positive. From (23) it is clear that the earlier requirement of large values of s/b, which make the asymptotic equation (11) valid, must be modified by the new requirement to keep w/T small. The clash is only serious for small specimens of soft materials: for these the measurements may be restricted to relatively small values of s/b, and the plot of reduced variables must then be calibrated indirectly, by comparison with a case in which the last factor in (23) is sufficiently small.

Besides the error due to Q, which is bound up with the principle of the experiment, there are also the errors due to the various imperfections of the apparatus. Most of these, as for instance the errors in the measurement of δ and s, are easy to assess and need not be discussed. The error produced by deviations of the crack from the centre is more important, and it may be useful to give an estimate of this.

For given δ the strain energy $U(s, \delta)$ is a maximum if the crack runs in the centre. If the crack divides the specimen into beams of unequal height, say ξ and $b-\xi$, then they will take up unequal parts of the total separation δ, in such a way that the sum of the strain energies is minimized. To simplify the equations, the special values $b = \delta = 2$ will be considered. That part of δ which is taken up by the beam of height ξ will be denoted by α. Suppressing irrelevant constants, the total strain energy can be written $U = \xi^3\alpha^2 + (2-\xi)^3(2-\alpha)^2$. Taking $\xi = 1+\zeta$ and $\alpha = 1+\beta$ brings in the deviations from the centre. Substitution then gives for the terms up to second order $U = 2 + 6\zeta^2 + 12\zeta\beta + 2\beta^2$. From this, $dU/d\beta = 12\zeta + 4\beta$, and the minimum is at $\beta = -3\zeta$. Hence the energy is $U = 2 - 12\zeta^2$ or $U/U_0 = 1 - 6\zeta^2$, where U_0 is the energy for the central crack. Since ζ measures the deviation in units of the half-width, the error in U will not exceed 6% so long as the crack remains within the central tenth of the specimen. This was the criterion adopted during the experiments for accepting measurements of δ and s. The error in T as evaluated from (11) is again positive. If the distribution of cracks within the allowed domain was uniform, the average error in T from this cause would be $+2\%$.

Both the systematic error due to Q and that due to non-central cracks can be diminished substantially by setting stricter standards for the data. In the present experiments the scatter of the results implies an uncertainty of about $\pm 10\%$ in T, so that the total systematic error of perhaps $+3\%$ is not significant.

§ 5. EXPERIMENTAL RESULTS

The elastic and many other properties of polymers depend on the time scale of the experiment, and also on the temperature. This makes it necessary, in the present case, to specify the crack speed; the temperature was not controlled and was in the region between 15°c and 20°c. Most measurements were made on slow cracks, in accordance with the original plan. It was later found that the apparatus can also produce fast cracks; observations on these will be discussed further below.

The slow cracks could be controlled smoothly by movement of the wedge, and their speed of propagation was from zero up to about $0{\cdot}1\ \mathrm{cm\,sec^{-1}}$. Readings were taken when the crack was stationary: this means a speed less than $0{\cdot}001\ \mathrm{cm\,sec^{-1}}$. The whole experiment with one specimen took about 15 minutes. The value of E for a loading time of a few minutes was determined separately, by a bending experiment on a cantilever beam cut from the same sheet as the specimens. The results were $E = 2{\cdot}78 \times 10^{10}\ \mathrm{dyn\,cm^{-2}}$ for the polymethylmethacrylate sample, and $E = 3{\cdot}07 \times 10^{10}\ \mathrm{dyn\,cm^{-2}}$ for the polystyrene. These agree well with published values for similar loading times at room temperature.

The results of the measurements on the slow cracks are shown in figure 3, in the form of a plot of $(\delta^2/s)^{1/3}$ against s/b. It is not practicable in the small-scale figure to distinguish the specimens of varying size by showing individual points.

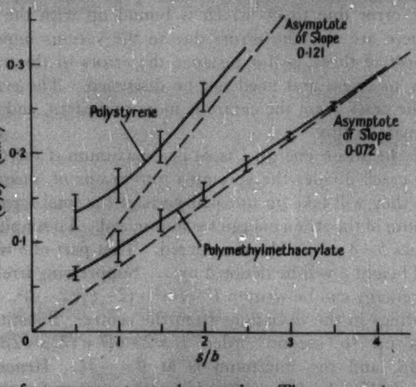


Figure 3. Results of measurements on slow cracks. The curves give mean values and the vertical bars the outside limits of the scatter.

Within the scatter, for which the outside limits are shown, the results superpose for all specimens, and no trend with b could be detected. This means that the scaling law (1) is, at least, a good approximation. More precisely, for a variation of the characteristic stress in the ratio 1:2, the specific fracture energy of polymethylmethacrylate did not vary by more than 1:1·1. With polystyrene the characteristic stress was only varied in the ratio $1:\sqrt{2}$; the fracture energy again did not vary by more than 1:1·1.

The theoretical asymptote (11) is a straight line through the origin of the plot. Its slope is the quantity $(64T/3E)^{1/3}$. The figure shows how the empirical curves tend towards straight lines through the origin as s/b increases. The form of the approach is determined by the universal function $\phi(b/s)$ in equation (1), and is, therefore, the same for any substance. This justifies the drawing of the asymptote for polystyrene, despite the limited range of the experiments. Measurements for larger s/b were not feasible for polystyrene, since the error given by equation (23) would have become too large.

For polymethylmethacrylate the slope of the asymptote is $(0{\cdot}072 \pm 0{\cdot}002)\,\mathrm{cm}^{-1/3}$; with the value of E given above, this results in

$$T = (4{\cdot}9 \pm 0{\cdot}5) \times 10^5 \,\mathrm{dyn\,cm}^{-1}.$$

For polystyrene the slope is $(0{\cdot}121 \pm 0{\cdot}003)\,\mathrm{cm}^{-1/3}$; this results in

$$T = (25{\cdot}5 \pm 3) \times 10^5 \,\mathrm{dyn\,cm}^{-1}.$$

The cracks in polymethylmethacrylate are usually mirror-smooth, except for faint marks which show where the crack edge was stationary. In polystyrene the surface of the cracks is rough, and the edge of the slow cracks is blunt. Some other materials were tried, to see whether stable cracks could be made. With a type of commercial asbestos sheet this proved impossible: the material failed by yielding, so that the beams sheared off near the end of the saw-cut. A type of hard phenolic resin gave straight cracks on which measurements were possible. Not unexpectedly, sheets of hardwood (oak) gave straight cracks along the grain even when the compression was not applied, but the end of the crack was difficult to locate. This difficulty is likely to arise whenever the material is not homogeneous.

Fast cracks were first observed in the present experiments when various liquids were applied to the specimens. It was thought that contact with solvents might lower the fracture energy. For a test, cracks in polymethylmethacrylate were filled with ethyl alcohol, carbon tetrachloride or petroleum ether. The result was the opposite to that expected: the wet cracks require more energy to grow; apparently the solvent, by softening the material around the crack edge, causes additional plastic flow. The sequence of events in the experiment is as follows: for a time, while δ is gradually increased by means of the wedge, the crack does not grow at all; eventually it starts off quickly and runs into the dry material ahead. For polymethylmethacrylate, the crack length attained in the jump is equal to the length of a slow crack in dry material, at equal values of δ. Jumps can also be produced by starting the crack from a blunt edge formed in dry material, for instance by omitting the swallow-tail on the initial saw-cut and by using a thick blade.

With polystyrene the crack can be made to jump, at any stage of the experiment, by advancing the wedge quickly. In fact, with the widest specimens ($b = 4$ inches) the crack jumps so easily that great care is necessary to obtain slow cracks. In polystyrene the length attained by a jumping crack is far greater than the length of a slow crack for the same value of δ. Measurements of this effect were made as follows. The wedge is driven in until a jump occurs; after the jump the variables have the values s_1, δ_1, which are measured. Subsequently, δ is increased slowly, until the crack begins to propagate (it may jump again); at this stage δ has the value δ_2. By this procedure δ_1 and δ_2 are both associated with the propagation of a crack of length s_1 and with a stress field of fixed geometry, but δ_1 refers to a fast crack and δ_2 to a slow or slowly starting crack. If the analysis given in the earlier sections applies, the ratio of the specific fracture energies is $T_{\text{slow}}/T_{\text{fast}}=(\delta_2/\delta_1)^2$. The ratio can thus be obtained without determining the individual values of T_{slow} and T_{fast} separately. This was done for the phenolic resin, mentioned earlier, which also gave jumping cracks, and where the measurements were not extensive enough to determine T itself. The phenolic resin gave $(\delta_2/\delta_1)^2=2$. With polystyrene the jumps were varied systematically by choosing the start at different values of s. The results for the energy ratio were consistent and the average was $(\delta_2/\delta_1)^2=9$, which implies a remarkably large change in T. The changes in the fracture energy of polymethylmethacrylate caused by the solvents could also be evaluated in terms of $(\delta_2/\delta_1)^2$. The increase amounts to a factor 2·5 for petroleum ether, a factor 3·0 for C_2H_5OH, and a factor 3·2 for CCl_4.

There is a difficulty about the interpretation of the jumps, because all the analysis in the present paper applies to quasi-static conditions only. For this reason, the equation $(\delta_2/\delta_1)^2=T_{\text{slow}}/T_{\text{fast}}$ is only well founded so long as the speed of the fast crack remains small compared with the velocity of elastic waves. As there is no evidence on this point, the meaning of the results obtained with fast cracks cannot be regarded as definitely established.

§ 6. SUPPLEMENTARY DISCUSSION

The specific fracture energy of polymethylmethacrylate has been measured earlier by Kies (1953). The average of his results is

$$T=3{\cdot}5\ \text{lb in}^{-1}=6{\cdot}1\times10^5\ \text{dyn cm}^{-1}.$$

This is 25% higher than the present value. Considering that the sample of material was not the same, and that the conditions of the test were different, the disagreement is not significant.

The qualitative difference between the behaviour of polymethylmethacrylate and that of polystyrene is surprising. It is opposite to what might be predicted from the known visco-elastic properties of these materials. When tested in oscillatory experiments at small strains, the damping is much less for polystyrene, and as a consequence its modulus rises more slowly with increasing frequency. Yet in the cracking experiments it is polystyrene which shows a marked increase in brittleness for fast cracks. Looking back, other signs of this paradox are apparent. Polystyrene is ductile and soft when worked by drilling or sawing, and the surface is easily scratched; but the shatter resistance of polystyrene, as tested by a hammer blow against a sheet, is not very high. The effects observed with the jumping cracks now give a more quantitative measure for this contrast of properties.

An increase of brittleness at higher strain rates is a well-known effect with metals. It is interesting to speculate whether the type of apparatus described here could be applied to metals. The relevant equations are (12) and (23). The latter shows that for metals the error caused by Q would be smaller than for the plastics considered here. Taking steel, for instance, E can be expected to be larger by a factor 100, whereas T will only be larger by a factor 10. Thus for specimens of the same size the last factor in equation (23) will decrease. Equation (12) shows, however, that in experiments with metals the forces would be large. For steel, using the same figures as above, the transverse force P will increase by a factor 30, in comparison with that for plastics. Now for the plastics a typical value of P is 50 kg, with the specimens used; thus for steel and similar specimens the force P would be 1·5 tons, which requires a stronger apparatus.

There is some indirect evidence that the stabilization of the crack by lengthwise compression will work with metals. Robertson and Hunt (1953) describe experiments on the splitting of short steel bars by hammer blows against one end; their test piece is fabricated in a special form shown in their paper. By a mechanism peculiar to this form, the hammering produces transverse forces. Cracks could thus be propagated downwards into the test piece. Naturally, however, the hammering will also produce compression along the test piece. This circumstance was possibly essential for the success of their experiment.

REFERENCES

GRIFFITH, A. A., 1920, *Phil. Trans. Roy. Soc.*, **221,** 163; 1924, *Proc. Int. Congr. Appl. Mech.* (*Delft*), 55.

KIES, J. A., 1953, *U.S. Naval Res. Lab. Report*, No. 237.

OBREIMOFF, J. W., 1930, *Proc. Roy. Soc.* A, **127,** 290.

RIVLIN, R. S., and THOMAS, A. G., 1953, *J. Polymer Sci.*, **10,** 291.

ROBERTSON, T. S., and HUNT, D. LE M., 1953, *Conference on Brittle Fracture of Steel*, The West of Scotland Iron and Steel Institute, Paper No. 456.

ROESLER, F. C., 1956, *Proc. Phys. Soc.* B, **69,** 981.

High Power Discharges in an Intense Longitudinal Magnetic Field†

By K. S. W. CHAMPION
Electron Physics Department, University of Birmingham‡

Communicated by J. Sayers; MS. received 9th March 1955, *and in revised form* 10*th September* 1956

Abstract. The properties of high power pulsed discharges in hydrogen have been studied. The time for the discharge to reach steady state conditions was investigated as a function of its parameters. Without a magnetic field the time was of the order of 100 μsec. At high pressures, longitudinal magnetic fields up to 23 000 oersteds caused little change in the time to reach equilibrium, but at low pressures (e.g. 0·46 mm Hg) the time was considerably increased and values up to 2·5 msec were recorded. Observations of the arc diameter showed that a strong uniform magnetic field produced a constriction, unless the natural arc diameter was much greater than the tube diameter. Non-uniform magnetic fields, on the other hand, caused the discharge to become diffuse where the field was diverging.

The electrode falls and positive column gradient of the steady state discharges were measured. Without a magnetic field the electrode falls were 58 v, and the positive column gradient was $60p^{0\cdot34}\,i^{-0\cdot6}$ where p is the pressure in mm Hg and i the current in amperes. This expression is in reasonable agreement with the theoretical value. The electron densities were found to lie in the range 10^{13} to 10^{15} electrons per cm^3. These densities correspond to ionization of from 0·5% to 12% of the gas atoms. The addition of a magnetic field considerably increased the electrode falls and values as high as 400 v were recorded. At low pressures a magnetic field, as predicted by theory, reduced the positive column gradient. At high pressures, on the other hand, the gradient was increased. For example, with a 100 A arc at 10 mm pressure the gradient was increased from 13 v per cm to 25 v per cm by an intense magnetic field. Again this change is consistent with theory. This change in gradient was accompanied by a 13% increase in the discharge temperature.

§ 1. INTRODUCTION

For some years there has been considerable uncertainty as to whether or not a uniform longitudinal magnetic field constricts a gas discharge. Cummings and Tonks (1941) claimed that a uniform longitudinal field does not have this effect. Other research workers, however, including Reichrudel and Spiwak (1941) claim that a constriction occurs. The present work shows that if appropriate boundary conditions are satisfied a constriction occurs, but if these conditions are not satisfied no constriction is observed.

In addition, some of the physical properties of high power discharges in intense longitudinal magnetic fields have been studied for the first time. Currents of about 100 A were passed through hydrogen at low pressure and magnetic fields of from zero to 23 000 oersteds were applied. Massey *et al.* (1949) investigated low pressure discharges in magnetic fields up to 11 400 oersteds, but the currents used did not exceed 3 A and in any case the experimental conditions were such that no positive column but only a negative glow was obtained.

---

† The material contained in this paper is based on portion of a Ph.D. thesis submitted to the University of Birmingham (1951).

‡ Now at the Department of Physics, Tufts University, Medford, Massachusetts.

§ 2. EXPERIMENTAL TECHNIQUE

Because of the very high power dissipation and the resultant cooling and gas clean-up problems both the discharge and the magnetic field were pulsed. The details of the method of production of the magnetic field and the electronic timing circuit for controlling the pulses have been described elsewhere (Champion 1950, Champion and Allen 1952). The discharge was produced by an approximately square, high power 2 msec pulse formed by triggering an artificial line. It was found that the time variation of the magnetic field was sufficiently slow for the discharge to adjust itself without any appreciable phase lag. This is consistent with the results of Kugler and Roggendorf (1933). However, for convenience, measurements of the steady state were made at the centre of the discharge pulse, phased as in figure 1. The repetition frequency of the pulses was usually one per second but, when desired, individual pulses could be produced.

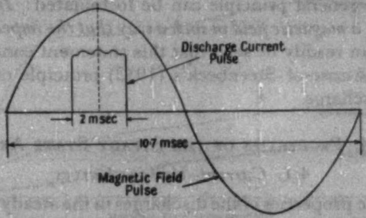


Figure 1. Magnetic field and arc current pulse wave forms showing their relative phases.

A number of different discharge tubes were used, including one with adjustable separation of the electrodes. The internal diameters of the tubes were approximately 4 cm and the electrode separations ranged from 5 to 28 cm. The tubes were normally placed centrally in the magnet coil so that the discharge lay on the axis of the coil. Because of the high energy in the discharge and consequent strong bombardment of the walls, it was necessary to outgas the tubes very thoroughly before making measurements. This was done first by baking in an oven and second by running the discharge itself, the tube being pumped out and refilled with hydrogen several times during the latter process. If the tubes were not adequately outgassed, the measured positive column gradients were lower by as much as a factor of two. This is, no doubt, due to the fact that the positive column gradient in hydrogen is substantially higher than that in any other gas.

§ 3. EFFECTS OF A LONGITUDINAL MAGNETIC FIELD

With a wide range of experimental conditions it was found that application of a uniform longitudinal magnetic field produced a constriction of the arc diameter. For example, a 40 A discharge at 0·95 mm Hg pressure was relatively tenuous, but application of a strong field caused it to be considerably constricted, with the formation of a narrow bright core about 4 mm in diameter. In general it was found that the magnetic field reduced the natural radius† of the arc and that if this was less than the radius of the containing tube a constriction was observed.

With a non-uniform longitudinal magnetic field which was strongest at the centre of the discharge and weaker (and hence diverging) near each electrode, it was found that the discharge diameter at the centre was approximately the same as when in a uniform field of the same strength. Low pressure discharges were less constricted near the electrodes. With high pressures (greater than 2 mm Hg) where the discharge does not fill the tube cross section, the application of the non-uniform field actually made the discharge more diffuse near the electrodes than it was without the field. This effect was greater near the anode than near the cathode and was probably due to the diverging field being more effective in deflecting electrons approaching the anode than ions approaching the cathode.

It could be deduced that in the above experiment when electrons had to cross lines of magnetic force they tended to do so where the field was weakest (i.e. near the electrodes). By combining this conclusion with results obtained with discharges transverse to a magnetic field (Somerville, Champion and Bigg 1948) the following general principle can be formulated: *In a gas discharge the electrons tend to cross a magnetic field in such a way that the impedance of the discharge is a minimum.* It can readily be seen that this statement constitutes an extension to the magnetic field case of Steenbeck's (1932) principle of minimum running voltage for a gas discharge.

§ 4. PROPERTIES OF THE STEADY STATE ARCS

4.1. *Current–Voltage Curves*

In this section the properties of the discharges in the steady state are considered, the criterion for the steady state being that the potential across the discharge remains constant as a function of time when all the operating conditions are maintained constant. Measurements were made in the pressure range 0·4 to 10 mm Hg.

In figures 2 and 3 typical current–voltage curves are plotted. Without a magnetic field the curve is almost a straight line with a small negative gradient. From figure 3 it can be seen that with a magnetic field the shape of the curves at 10 mm Hg pressure is little changed, but the arc voltage is steadily increased. The second mode was unusual and was shown to be a high current glow. The curves in figure 2 are typical of those obtained at 0·46 mm Hg and with higher pressures they gradually changed to those obtained at 10 mm Hg. The curves at low pressure are characterized by turning up sharply at low currents, this tendency increasing with magnetic field strength.

A more useful way to represent the information contained in the current–voltage curves is to plot the 'effective arc impedance' as a function of magnetic field strength. This leaves arc length, current and pressure as parameters.

---

† The natural radius of an arc (cf. Champion 1952 a) is the radius the arc assumes if not restricted by a containing tube. If the natural radius changes but is still greater than the tube radius there is no change in the density distribution inside the containing tube.

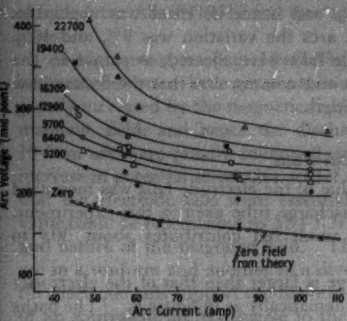


Figure 2. Current–voltage curves (pulse mid-point) for arc 19·25 cm long in 0·46 mm Hg of hydrogen.

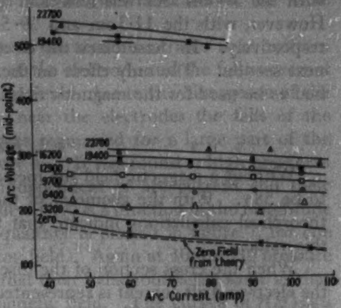


Figure 3. Current–voltage curves (pulse mid-point) for arc 5·4 cm long in 10 mm Hg of hydrogen.

Values of field given on curves are in oersteds.

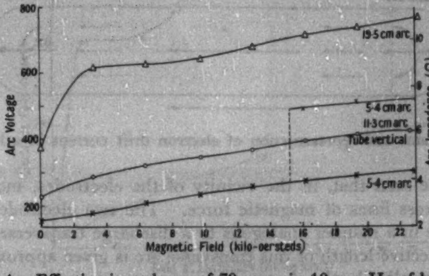


Figure 4. Effective impedance of 70 A arcs in 10 mm Hg of hydrogen.

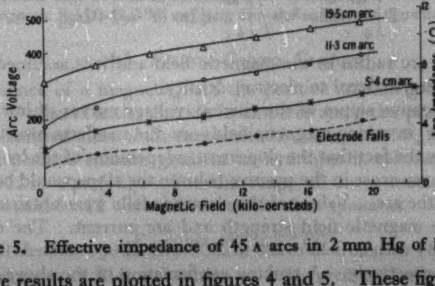


Figure 5. Effective impedance of 45 A arcs in 2 mm Hg of hydrogen.

Representative results are plotted in figures 4 and 5. These figures correspond to specified arc currents and pressures and thus show the dependence of the impedance on arc length.

The magnetic field used in these experiments was cylindrically symmetrical, but was weaker near the ends of the long arcs than at their centre. The variation with the 5·4 cm arcs was less than 1·8%, and hence its effect was negligible. However, with the 11·3 cm and 19·5 cm arcs the variation was 9% and 40% respectively. In these cases the electrode falls were affected, as shown in the next section. The only effect on the positive column was that the mean value had to be used for the magnetic field strength.

4.2. *Electrode Falls*

Without a magnetic field the mean value of the electrode falls was measured to be 58 v. With the geometry of the discharge tube used in this experiment, theoretical calculations indicate that the anode fall contributes about 30 v to this value.

When the cross section of the discharge is larger than that of the electrodes the electron drift current is represented schematically as in figure 6. (The paths of individual electrons do not follow these lines.) When a longitudinal magnetic field is applied the section between the dotted lines constitutes the positive column of an arc parallel to the field and the other two sections constitute the electrode falls. It can be seen that, in the vicinity of the electrodes, most of the drift current must cross lines of magnetic force. The two electrode regions taken together behave in a manner analogous to a discharge transverse to a magnetic field.† The effective length of this transverse arc is given approximately by the mean distance travelled by the electron current across the magnetic field. This is

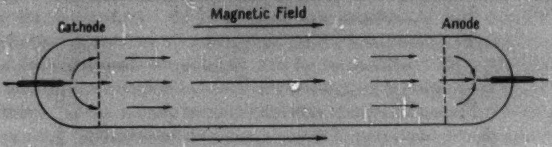


Figure 6. Diagrammatic representation of electron drift current in a low pressure arc.

$$2\int_0^{r_{aH}} n_e 2\pi r^2\,dr \Big/ \int_0^{r_{aH}} n_e 2\pi r\,dr = 1{\cdot}02 r_{aH} \qquad \ldots\ldots(1)$$

where r_{aH} is the arc radius in the magnetic field and it is assumed that the rate of ionization is proportional to electron density.

The large negative slopes of the current–voltage curves of low pressure arcs (e.g. 0·46 mm Hg) in strong magnetic fields are due to effects near the electrodes. This is shown by the fact that the slopes are independent of the length of the arc, whereas if the cause arose in the positive column the slope would be proportional to the length of the arc. Values of the electrode falls were obtained which were functions of the magnetic field strength and arc current. The electrode falls increased from 58 v with no field to 200 to 250 v with 22 000 oersteds for arc currents of 100 to 40 A respectively. A striking confirmation of the above hypothesis as to the nature of the electrode falls was obtained from curves showing their dependence on current and field strength, which were very similar to those obtained by Allen (1951) for the voltage across transverse arcs at the same pressure.

---

† With a magnetic field diverging near the electrodes this picture still applies essentially, as the electron drift current still crosses the magnetic lines near the electrodes.

At high pressures (e.g. 10 mm Hg) the cross sections of arcs in a uniform magnetic field were little greater than that of the electrode and the falls were not greatly increased by the magnetic field. However, due to the divergence of the magnetic field, and hence the discharge, near the electrodes the falls of the 19·5 cm arc were greatly increased, and this accounted for a large part of the increase in potential across this arc (cf. figure 4). It was found at high pressures that the electrode falls were essentially independent of arc current and their value could be represented by $58+f(H)$ where $f(H)$ is a function of the magnitude and nature of the magnetic field. This quantity is plotted in figure 7, for one arc in a uniform and another in a diverging field. Again at 10mm Hg pressure the electrode falls support the hypothesis that their behaviour is the same as that of an equivalent arc transverse to the magnetic field. The curves of $f(H)$ as a function of magnetic field are exactly the same as those which Allen (1951, figures 1 and 2) obtained with transverse arcs under corresponding conditions.

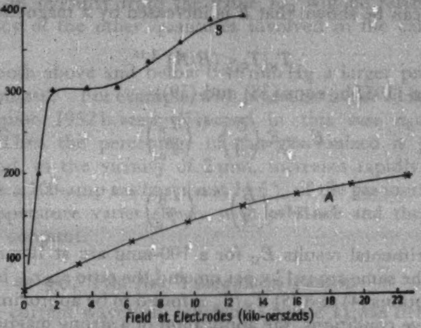


Figure 7. Electrode falls in arcs in 10 mm Hg of hydrogen. A, 5·4 cm long arc (uniform magnetic field); B, 19·5 cm long arc (non-uniform magnetic field).

4.3. *Positive Column Gradient*

In the absence of a magnetic field the positive column gradient was found to be $60p^{0\cdot34}i^{-0\cdot5}$ where i is the current in amperes and p the pressure in mm Hg. This expression fitted most of the experimental data within 5%. The theoretical value for the positive column gradient in hydrogen is calculated below (from Champion 1952 b, eqn (12)). For the range of the present experiments the values $\nu=0\cdot04$ and $\delta=12$ were substituted in the exponents. This gives

$$E=\text{const.}\,p^{0\cdot104}i^{-0\cdot89}T^{0\cdot76} \qquad \ldots\ldots(2)$$

where the temperature T is a function of both pressure and current. These functions are not known accurately, but over the pressure range 1 to 10 mm Hg and currents in the region of 100 A calculations based on Wasserrab's (1951) theory of arc temperatures yield $T\propto p^{0\cdot25}i^{0\cdot5}$. This gives $E=\text{const.}\,p^{0\cdot29}i^{-0\cdot51}$ which is in reasonable agreement with the experimental value.

Theory predicts that at low pressures the positive column gradient will be reduced by a longitudinal magnetic field. Experimentally it was found that, with 0·46 mm Hg pressure, E was reduced by the magnetic field over the whole current range investigated (40 to 110 A). With values of E measured 1 msec after the beginning of the pulse the reduction was a maximum with a field of approximately 11 000 oersteds and the value of E increased again with stronger fields until with 23 000 oersteds its value was nearly as large as its no field value. However, as pointed out in § 5 the formation period was much longer with these conditions and times up to 2·5 msec were measured. Thus the fact that the plasma had not completely reached equilibrium conditions probably accounted for the apparently smaller reduction in E with very strong magnetic fields.

The value of 2 mm Hg approximately marked the transition to high pressure conditions in the present experiments and at this pressure the magnetic field caused little change in the positive column gradient. At 10 mm Hg it was found experimentally that the magnetic field increased the value of E. For these arcs the value of ν, the exponent in the Nusselt relation, and δ, the exponent giving the temperature dependence of electron density, are little affected by a magnetic field. Thus it can be shown that E is increased by a magnetic field (Champion 1952 b) unless

$$T_H/T_0<(B/A)^{-1\cdot15}. \qquad \ldots\ldots(3)$$

From Champion (1952 b, eqns (38) and (39))

$$E_H=E_0\left(\frac{B}{A}\right)^{0\cdot67}\left(\frac{T_H}{T_0}\right)^{0\cdot76} \qquad \ldots\ldots(4)$$

and

$$r_{aH}=r_{a0}\left(\frac{B}{A}\right)^{-0\cdot43}\left(\frac{T_H}{T_0}\right)^{-6\cdot72}. \qquad \ldots\ldots(5)$$

From the experimental results E_H for a 100-amp arc at 10 mm pressure is 25 v per cm, E_0 for the same arc is 13 v per cm and the ratio r_{a0}/r_{aH} is approximately 3. Thus from equations (4) and (5) T_H/T_0 is found to be approximately 1·13. That is, with the above conditions, the application of a strong magnetic field produces a 13% increase in arc temperature.

4.4. *Electron Density*

The electron density can be calculated from the appropriate arc current equation, viz.

$$i=0\cdot346\pi a^2en_0K_eE \qquad \ldots\ldots(6)$$

where a is the tube radius, n_0 the electron density at the axis and K_e the electron mobility. The numerical factor results from integrating the density distribution corresponding to ionization by stages. For a typical case at 0·46 mm Hg pressure with a current of 100 A, n_0 was found to be $3\cdot6\times10^{13}$ electrons per cm^3, where Bradbury and Nielsen's (1936) value was used for the mobility. The gas temperature was about 1000°K and thus the number of neutral atoms was $4\cdot4\times10^{15}$ per cm^3. Consequently, in this arc about 0·8% of the gas was ionized.

At this point it may be appropriate to justify the use of Bradbury and Nielsen's value for the mobility under these conditions. In their experiment the electron mobility in H_2 was measured. In the present experiments the electrons move in a gas of H_2, H and H^+ (possibly also H_2^+ and H_3^+). The effect of H^+ on the mobility as measured by Bradbury and Nielsen (K_B) will first be considered:

$$v_d = E\left(\frac{1}{K_B} + \frac{1}{K_C}\right)^{-1} \qquad \ldots\ldots(7)$$

where v_d is the drift velocity and K_C the mobility if the electrons and ions alone were present.

$$K_B = \frac{3{\cdot}7 \times 10^5}{p} = 8 \times 10^5 \text{ cm sec}^{-1}/\text{v cm}^{-1} \text{ when } p = 0{\cdot}46 \text{ mm Hg}$$

$$K_C = \frac{\sigma}{n_0 e} = \frac{40}{3{\cdot}6 \times 10^{13} \times 1{\cdot}6 \times 10^{-19}} = 7 \times 10^6 \text{ cm sec}^{-1}/\text{v cm}^{-1}.$$

The value $\sigma = 40$ mho cm^{-1} is quoted from H. Dreicer (private communication, 1954). Thus, in the case considered, the effect of H^+ will be to reduce the mobility by about 10%. Accurate cross sections for H are not available. However, there is probably not a large proportion of H in this particular arc and in any case, the cross section of two H atoms is approximately the same as that of an H_2 molecule for elastic collisions of electrons of not more than 4 ev energy. Thus any change in mobility from the value K_B will be relatively small and within the accuracy of the other quantities involved in the calculation of the electron density.

At pressures both above and below 0·46 mm Hg a larger proportion of the gas in the tube is ionized. For example, with pressures in the vicinity of 2 mm Hg, $n_0 \propto p^{0{\cdot}5} T^{\delta}$ (Champion 1952 b, eqn (5)) and in this case $n_0 \propto p^{1{\cdot}63}$, whereas $n_g \propto p/T \propto p^{0{\cdot}75}$. Thus the percentage of the gas ionized is proportional to $p^{1{\cdot}63}/p^{0{\cdot}75} \propto p^{0{\cdot}88}$ and, in the vicinity of 2 mm, increases rapidly with pressure. At 10 mm pressure a 100-amp arc has about 11·5% of the gas ionized. At higher pressures the temperature varies slowly with pressure and the value of n_0/n_g remains relatively constant.

§ 5. THE ARC FORMATION PERIOD

The discharge current waveform (shown in figure 1) was determined by the characteristics of the artificial line and 7% ripple was present. In most cases the arc current–voltage curves had nearly zero slope and thus the ripple in the voltage waveforms was relatively small.

In figure 8 are shown typical voltage waveforms, obtained without and with a longitudinal magnetic field. In the first trace can be seen the sharp initial voltage peak (AP) corresponding to the spark which initiates the discharge. This was followed by an interval, usually about 100 μsec long, during which there was a relatively slow and small drop in voltage, merging into the equilibrium region. The equilibrium discharges in almost all cases had the properties of arcs. However, with lower currents (e.g. 4 A) it was possible to obtain glow discharges, with cathode falls of the order of 200 v (with no magnetic field). Because of their high impedance, the voltage waveform of the glows was primarily determined by the artificial line. With slightly higher currents it was possible to obtain discharges which changed to arcs part way through the pulse, as shown in figure 2(*c*). In this case, when a magnetic field was applied, the discharge was unstable, making rapid transitions from one state to the other. This instability seems to be associated with the effect of the intense magnetic field on the motion of the charged particles near the cathode.

The noise shown on the voltage waveforms in the presence of the magnetic field was invariably observed when a magnetic field was applied, and is due to oscillations caused by the magnetic field making the plasma non-isotropic.

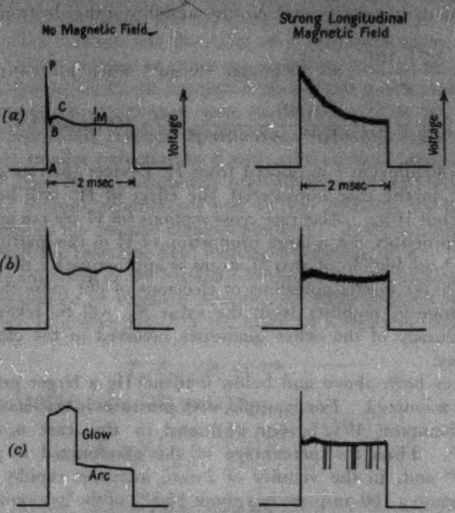


Figure 8. Voltage waveforms (*a*) with 0·46 mm Hg of hydrogen, (*b*) with 10 mm Hg of hydrogen (11·5 cm long arc), (*c*) showing glow-arc transitions.

5.1. *With No Magnetic Field*

The arc formation period corresponds to the transition from the quasi-equilibrium spark conditions to the equilibrium arc conditions at point M (figure 8 (*a*)) and is characterized by point C. Values of the voltage difference AC–AM were measured, ranging from 5 v to 50 v. Analysis of the results showed that a portion of this voltage arose in the positive column, but also that 5 v to 20 v was due to electrode (probably cathode) effects. The latter voltage was shown to be dependent on the electrode surfaces by changing electrodes and noting the change in its magnitude. This observation is in agreement with Haworth (1950) who found that the initial arc potential (AC in the present notation) was variable and seemed to be determined by the nature or presence of an extremely thin insulating layer on the cathode surface.

The positive column contribution to the voltage difference AC–AM is due to the time taken for the neutral and charged particles in the positive column to reach their final density and energy values. At low pressures (e.g. 0·46 mm Hg) the mean positive ion drift velocity is $7{\cdot}0 \times 10^4\,\mathrm{cm\,sec^{-1}}$. This gives a transit time of 0·28 msec for the ions in a 19·5 cm arc. In this case the time to reach complete equilibrium is probably related to the ion transit time. The magnitude of AC – AM was smaller when the initial rate of current rise was reduced, since the discharge had more time to adjust itself to the conditions appropriate to the large current.

The cathode emission in an equilibrium arc is probably of a modified thermionic type. Thus the time taken for a portion of the cathode to be heated to a temperature at which it has copious thermionic emission is of interest. During this time additional energy is absorbed by the cathode and this requires a slightly larger cathode potential fall. Llewellyn Jones (1950)† gives an approximate expression for the time required to raise the hot-spot temperature to the boiling point of the metal. By assuming that all the incident energy W is used to raise the temperature of the metal it can be shown that

$$T_c - T_0 = \frac{W}{J}\left(\frac{t}{\rho s k}\right)^{1/2} \qquad \ldots\ldots(8)$$

where T_c is the spot temperature, T_0 the initial temperature of the metal, t the time during which the cathode has been heated, ρ the density, s the specific heat and k the thermal conductivity of the cathode material. Equation (8) holds when

$$\left.\begin{array}{l} l\left(\dfrac{\rho s}{kt}\right) \gg 1 \\ \\ \text{and } l \leqslant r_s \end{array}\right\} \qquad \ldots\ldots(9)$$

where l is the depth and r_s the radius of the cathode spot. If the first of these conditions is not satisfied there is another term in the integral from which is obtained equation (8). If the second condition is not satisfied, equation (8) is not accurate since lateral heat losses from the spot have been neglected in its derivation.

In the present experiment equation (8) gives a time 66 μsec for the surface of the cathode spot to reach boiling point. From equation (9) $l \gg 6{\cdot}4 \times 10^{-3}$ cm, say $l \sim 1{\cdot}2 \times 10^{-2}$ cm, but $r_s \sim 5{\cdot}7 \times 10^{-3}$ cm. This means that $l > r_s$ and that the lateral flow of heat is not negligible. If this additional loss is corrected for, the time becomes approximately 120 μsec and this value may be further increased by radiation losses. Thus the origin of the cathode contributions to the voltage difference AC − AM can be seen.

5.2. *In the Presence of a Magnetic Field*

At higher pressures (2 mm Hg and above) the application of a magnetic field caused little change in the pulse waveform, except for the superposition of a considerable amount of noise. The formation period was only very slightly increased. This was due to the high pressure (and hence collision frequency) making motion of the charged particles across the magnetic field relatively easy.

At low pressures (e.g. 0·46 mm Hg) the arc cross section was considerably greater than that of the electrodes and the particle collision frequency was lower. Thus, application of a magnetic field caused considerable change in the pulse waveform and the formation period was greatly increased. The formation period was found to be approximately proportional to the length of the arc. In figure 8 (*a*) is sketched a typical waveform, obtained with an arc 11·5 cm long in the presence of a strong magnetic field. In this case the formation period was 1·2 msec.

---

† For more recent studies of this subject see Somerville, Blevin and Fletcher (1952).

ACKNOWLEDGMENT

The author wishes to thank Professor J. Sayers for suggesting the project of which the above work formed part.

REFERENCES

ALLEN, N. L., 1951, *Proc. Phys. Soc.* B, **64**, 276.
BRADBURY, N. E., and NIELSEN, R. A., 1936, *Phys. Rev.*, **49**, 388.
CHAMPION, K. S. W., 1950, *Proc. Phys. Soc.* B, **63**, 795; 1952 a, *Ibid.*, **65**, 329; 1952 b, *Ibid.*, **65**, 345.
CHAMPION, K. S. W., and ALLEN, N. L., 1952, *Electron. Engng*, Sept., 423.
CUMMINGS, C. S., and TONKS, L., 1941, *Phys. Rev.*, **59**, 514.
HAWORTH, F. E., 1950, *Phys. Rev.*, **80**, 223.
KUGLER, G. A., and ROGGENDORF, A., 1933, *Phys. Z.*, **34**, 136.
LLEWELLYN JONES, F., 1950, *Brit. J. App. Phys.*, **1**, 60.
MASSEY, H. S. W., *et al.*, 1949, *The Characteristics of Electrical Discharges in Magnetic Fields* (New York : McGraw-Hill).
REICHRUDEL, E., and SPIWAK, G., 1941, *J. Phys., Moscow*, **4**, 211.
SOMERVILLE, J. M., BLEVIN, W. R., and FLETCHER, N. H., 1952, *Proc. Phys. Soc.* B, **65**, 963.
SOMERVILLE, J. M., CHAMPION, K. S. W., and BIGG, E. K., 1948, *Aust. J. Sci. Res.* A, **1**, 400.
STEENBECK, M., 1932, *Phys. Z.*, **33**, 809.
WASSERRAB, T., 1951, *Z. Phys.*, **129**, 190.

Drift of Minority Carriers in the Presence of Trapping

By A. K. JONSCHER

Research Laboratories, The General Electric Company, Ltd., Wembley, Middx.

MS. received 23rd August 1956, and in final form 8th November 1956

Abstract. A simple model of a semiconductor with one trapping level in equilibrium with one band only is defined through the lifetime τ and two trapping time constants τ_c and τ_s. A rectangular pulse of minority carriers is injected at one end and the drift of this pulse in a constant electric field is studied, neglecting diffusion effects. Analytical expressions are derived for the densities of free carriers N and of trapped carriers N_t, as functions of distance and time. Graphs are plotted of the time dependence of N at a collector point, corresponding to drift-decay measurements, and of the spatial distribution of N and N_t at different times, which are relevant in photoconductive measurements. It is concluded that the attenuation of the drifting pulse corresponds to an apparent lifetime which is shorter than the true lifetime τ and that the shape of the (N, time response) curve depends critically not only on the parameters τ, τ_c and τ_s but also on the distance of the collector point from the emitter and on the duration of the injecting pulse. In the case of heavy trapping most of the trapped and free charge is confined to the immediate neighbourhood of the emitter. In the steady state it is shown that, on this model, the density distribution of free carriers is entirely independent of trapping.

---

§ 1. Introduction

Attempts at explaining the observed rates of recombination of excess carriers in semiconductors, in particular in silicon and germanium, lead to the assumption of the presence of intermediate localized levels in the forbidden band which act as recombination centres. The simplest assumption is that these centres cause recombination of hole–electron pairs without changing their own occupancy by electrons. This implies that the excess densities of free electrons Δn and of free holes Δp are equal, in order to satisfy the electrical neutrality condition (Shockley and Read 1952)

$$\Delta n = \Delta p. \qquad \ldots\ldots(1)$$

We refer to this case as the 'no trapping' case.

Experiments with most semiconductors and photoconductors suggest the need for a more complicated model in which $\Delta n \neq \Delta p$, i.e. the population of some intermediate states changes as excess free carriers are introduced into the material. In general, we shall have

$$\Delta p = \Delta n + \Delta n_t = \Delta n - \Delta p_t, \qquad \ldots\ldots(2)$$

where Δn_t and Δp_t are the changes of concentration of electrons and holes respectively in the intermediate levels. The added concentration Δn_t is frequently referred to as trapped charge density.

Treatments of the transient response of models in which trapping is present, i.e. in which Δn_t is not zero, have been given for the case of uniform excitation throughout an infinite solid, that is in absence of carrier transport (Fan 1953, 1954, Haynes and Hornbeck 1955, Hornbeck and Haynes 1955 Rittner 1956). They show that the rise and decay of excess densities Δn and Δp on sudden application or removal of uniform generation do not, in general, follow a simple exponential law, as is the case in the absence of trapping. These results have a bearing on the interpretation of photoconductive measurements under conditions where the transport of carriers, and end effects, may be neglected.

A new situation arises when the problem of transport of carriers, i.e. drift and diffusion in the presence of trapping, has to be considered. The only published reference to this problem known to the author is a brief treatment by Fan (1953). The practical importance of these phenomena arises from the fact that drift in the presence of trapping is the basis of drift decay measurements of lifetime, while diffusion, both steady-state and a.c., is of importance in most semiconductor device applications.

This paper attempts to treat analytically the drift of carriers in a semiconductor in which a simple model for trapping has been assumed, similar to that used by Hornbeck and Haynes (1955). It is proposed to treat the steady-state and a.c. diffusion for the same model in a separate paper.

§ 2. THE PHYSICAL MODEL AND BASIC EQUATIONS

The physical model of a semiconductor exhibiting trapping effects adopted in the following analysis is shown in figure 1. A p-type semiconductor is assumed for definiteness, with a single trapping level of density N_1. The concentrations of electrons and holes are: $n=n_0+\Delta n$ electrons in the conduction band, $p=p_0+\Delta p$ holes in the valence band, $n_t=n_{t0}+\Delta n_t$ electrons in the trapping level and $p_t=p_{t0}-\Delta n_t$ holes in the trapping level. The subscript zero refers to thermal equilibrium densities and the Δ's denote added concentrations in excess of thermal equilibrium. We have the following obvious relation:

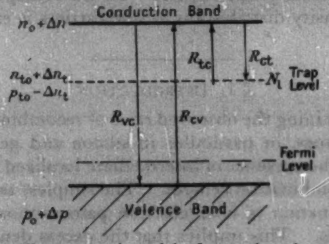


Figure 1. The assumed model of a semiconductor with traps.

$$n_t+p_t=n_{t0}+p_{t0}=N_1.$$

The level N_1 is postulated in equilibrium with the conduction band only through the transition rates R_{tc} and R_{ct}. A parallel and independent mechanism of recombination is assumed through the transition rates R_{cv} and R_{vc} giving a steady-state minority-carrier lifetime τ defined by

$$\Delta n_{\text{steady state}}=\tau g \qquad \ldots\ldots(3)$$

where g is the rate of generation in excess of the thermal rate. It is also assumed that, whatever states are responsible for the recombination process, their population does not change with Δn, so that $\Delta n = \Delta p$ for the recombination process alone.

The transition rates R_{ct} and R_{tc} may be expressed as follows:

$$R_{ct} = r_c n p_t = r_c (n_0 + \Delta n)(p_{t0} - \Delta n_t) \qquad \ldots\ldots(4)$$

$$R_{tc} = r_c \frac{n_0 p_{t0}}{n_{t0}} n_t = r_c \frac{n_0 p_{t0}}{n_{t0}} (n_{t0} + \Delta n_t) \qquad \ldots\ldots(5)$$

where r_c is a coefficient related to the transition probability. The rate of change of trapped density is

$$\frac{\partial \Delta n_t}{\partial t} = R_{ct} - R_{tc} = r_c \left(p_{t0} \Delta n - \frac{n_0 N_1}{n_{t0}} \Delta n_t - \Delta n \Delta n_t \right). \qquad \ldots\ldots(6)$$

In order to make this differential equation linear it is proposed to neglect the product term $\Delta n \Delta n_t$, i.e. to limit the range of validity of the analysis to low injection levels. We may then rewrite equation (6) in the form†

$$\frac{\partial \Delta n_t}{\partial t} = \frac{\Delta n}{\tau_s} - \frac{\Delta n_t}{\tau_c}, \qquad \ldots\ldots(7)$$

where $$1/\tau_s = r_c p_{t0} \qquad \ldots\ldots(8)$$

and $$1/\tau_c = r_c n_0 N_1 / n_{t0}. \qquad \ldots\ldots(9)$$

The continuity equation for free minority carriers is

$$\frac{\partial \Delta n}{\partial t} = -\frac{\Delta n}{\tau} + D_n \frac{\partial^2 \Delta n}{\partial x^2} + \mu_n E \frac{\partial \Delta n}{\partial x} - \frac{\partial \Delta n_t}{\partial t} + g \qquad \ldots\ldots(10)$$

where D_n and μ_n are diffusion constant and mobility for electrons.‡

It is proposed to neglect the diffusion of carriers and to consider a long filament (figure 2) with an injecting contact at one end, $x = 0$. A constant electric field $-E$

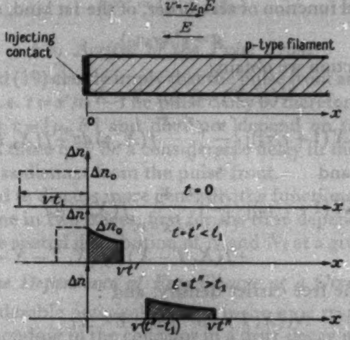


Figure 2. Boundary conditions for the injection and drift of a rectangular pulse of minority carriers.

---

† The following notations have been brought into line with the symbols used by Rittner (1956) and Fan (1953, 1954). The parameter r_c is identical with Fan's r_c, τ_s is analogous to Fan's and Rittner's τ_t, and τ_c is identical with their τ_c.

‡ Note that ambipolar diffusivity D^* and mobility μ^* should be used in the case of near-intrinsic material.

is applied along the specimen and surface effects are assumed to be accounted for through a suitable choice of τ, the steady-state lifetime for the filament. The injecting pulse is assumed to impress a constant excess density Δn_0 of minority carriers at $x=0$, starting with $t=0$ and extending to $t=t_1$.

Equations (7) and (10) are rewritten, using the following notations:

$$\left.\begin{aligned} N=\frac{\Delta n}{\Delta n_0},\ N_t=\frac{\Delta n_t}{\Delta n_0}, \\ T=\frac{t}{\tau},\ T_1=\frac{t_1}{\tau},\ X=\frac{x}{\lambda}=-\frac{x}{\mu_n E\tau}=\frac{x}{v\tau}, \end{aligned}\right\} \qquad \ldots\ldots(11)$$

where λ is the drift length and v the drift velocity. This gives:

$$\frac{\partial N}{\partial T}=-N-\frac{\partial N}{\partial X}-\frac{\partial N_t}{\partial T}, \qquad \ldots\ldots(12)$$

$$\frac{\partial N_t}{\partial T}=\frac{\tau}{\tau_s}N-\frac{\tau}{\tau_c}N_t. \qquad \ldots\ldots(13)$$

The boundary conditions are:

$$N=N_t=0 \text{ at } T=0 \text{ for all } X, \text{ and} \qquad \ldots\ldots(14)$$

$$N=1-\mathbf{1}(T_1) \text{ at } X=0, \qquad \ldots\ldots(15)$$

where $\mathbf{1}(T)$ is Heaviside's unit function.

§ 3. Solution of the Differential Equations

The solutions of equations (12) and (13) subject to the appropriate boundary conditions are shown below.

We define the function:

$$f(X,z)=\exp(-z\tau/\tau_c)\,\mathrm{J}_0[2i(aXz)^{1/2}],$$

where J_0 is the Bessel function of zero order, of the 1st kind, and

$$a=\tau^2/(\tau_s\tau_c).$$

In terms of this function we obtain:

$$N(X,T)=0 \text{ for } T<X, \qquad \ldots\ldots(16)$$

$$N(X,T)=\exp[-X(1+\tau/\tau_s)]\left\{\frac{\tau}{\tau_c}\int_0^{T-X} f(X,z)\,dz+f(X,T-X)\right\} \qquad \ldots\ldots(17)$$

for $X<T<T_1+X$, and

$$\begin{aligned} N(X,T)=\exp[-X(1+\tau/\tau_s)]\bigg\{\frac{\tau}{\tau_c}\int_{T-X-T_1}^{T-X} & f(X,z)\,dz \\ & +f(X,T-X)-f(X,T-X-T_1)\bigg\} \end{aligned} \qquad \ldots\ldots(18)$$

for $T>X+T_1$, for the free carrier density, and :

$$N_t(X,T)=0 \quad \text{for} \quad T<X, \qquad \ldots\ldots(19)$$

$$N_t(X,T)=\frac{\tau}{\tau_s}\exp[-X(1+\tau/\tau_s)]\int_0^{T-X} f(X,z)\,dz \qquad \ldots\ldots(20)$$

for $X<T<X+T_1$, and

$$N_t(X,T)=\frac{\tau}{\tau_s}\exp[-X(1+\tau/\tau_s)]\int_{T-X-T_1}^{T-X} f(X,z)\,dz \quad \text{for} \quad T>X+T_1, \qquad \ldots\ldots(21)$$

for the trapped carrier density.

§ 4. DISCUSSION OF ANALYTICAL SOLUTIONS

Before proceeding to a general discussion of equations (16) to (21), we note the following special points.

4.1. *Steady State for Continuous Injection*

In this case we put $T_1 \to \infty$, $T \to \infty$, and obtain, after some transformations:

$$N(X) = e^{-X}, \qquad \ldots\ldots(22)$$

$$N_t(X) = \frac{\tau_c}{\tau_s} e^{-X}. \qquad \ldots\ldots(23)$$

Thus the free and trapped charge densities decay with the proper steady-state minority-carrier lifetime τ, as if trapping was absent. The ratio $N_t/N = \tau_c/\tau_s$ everywhere.

4.2. *Response to a Very Short Pulse,* $T_1 \to 0$

Since $T - X \to 0$ and $T - X - T_1 \to 0$ in equations (17), (18), (20) and (21), the contribution of the integrals in these equations vanishes and the functions $f(X, T)$ cancel in (18) so that, in view of $f(X, 0) = 1$, we find:

$$\left.\begin{array}{l} N(X,T) = 0 \ \text{ for } \ T<X \ \text{ and } \ T>X, \\ N(X,T) = \exp\left[-X(1+\tau/\tau_s)\right] \ \text{ for } \ T = X, \end{array}\right\} \qquad \ldots\ldots(24)$$

while

$$N_t = 0 \ \text{ for all } X \text{ and } T. \qquad \ldots\ldots(25)$$

The implication is that a very short pulse is transmitted without distortion, since the traps have no time to respond fully, but the attenuation is determined by an apparent 'lifetime' τ^*, given by:

$$1/\tau^* = 1/\tau + 1/\tau_s. \qquad \ldots\ldots(26)$$

4.3. *Arrival of the Pulse Front*

Equations (16) and (19) clearly imply that the pulse front arrives at a given point X' in a time $T = X'$, i.e. $t = x'/v$. The pulse delay is therefore determined by the normal drift velocity $v = |\mu_n E|$ and does not depend on trapping. It will be shown, however, that there may be a considerable delay in the arrival of the maximum carrier density as distinct from the pulse front.

It is next proposed to discuss more generally the functional relations for N and N_t. This will be done in two stages, first for the time dependence of N at a fixed point and next for the spatial distribution of N and N_t at a given time.

4.4. *Time Dependence of Free Charge at a Given Point*

The case of considerable practical interest is the time response of N at a fixed point $X = X'$, corresponding to the collector in a drift-decay measurement. Here we may regard time as the only variable, and $X = X'$ as a parameter. Introducing the quantities

$$y = 4aX'T \quad \text{and} \quad m = 4\tau X'/\tau_s, \qquad \ldots\ldots(27)$$

so that $y = m\tau T/\tau_c$, and defining the function

$$\phi(y, m) = e^{-y/m}\, J_0(i\sqrt{y}), \qquad \ldots\ldots(28)$$

we may write

$$f(X', T-X') = \phi(y - m^2\tau_s/4\tau_c, m) \qquad \ldots\ldots(29)$$

and

$$\frac{\tau}{\tau_c}\int_0^T f(X', z)\,dz = \frac{1}{m}\int_0^y \phi(y, m)\,dy. \qquad \ldots\ldots(30)$$

It is easily shown that $\phi(y, m)$ is a monotonically decreasing function of y for $m<4$ and that it has a maximum if $m>4$. In the limit when $m \gg 4$, i.e. $X'\tau/\tau_s \gg 1$, the maximum occurs at $y \simeq m^2/4$ and the value of the function at the maximum is

$$\phi(m^2/4, m) \simeq e^{m/4}(\pi m)^{-1/2},$$

and thus the maximum value of $N(X', T)$ is found to be approximately

$$N_{\max}(X', T_m) = e^{-X'}\,T_1(\tau/\tau_c)(\pi m)^{-1/2}, \qquad \ldots\ldots(31)$$

occurring at approximately

$$T_m = X'(1 + \tau_c/\tau_s). \qquad \ldots\ldots(32)$$

This implies that when $m \gg 4$ the amplitude of $N(X', T)$ will continue to increase after the arrival of the pulse front (which may be indistinguishable on an oscillograph) and the peak of the pulse will correspond to an apparent mobility:

$$\mu' = \frac{\mu}{1 + \tau_c/\tau_s}. \qquad \ldots\ldots(33)$$

The case of $X'\tau/\tau_s > 1$ is shown in figure 3, giving the response as a function of time for three values of X' (1, 2 and 4) and for the given values of τ_c and τ_s.

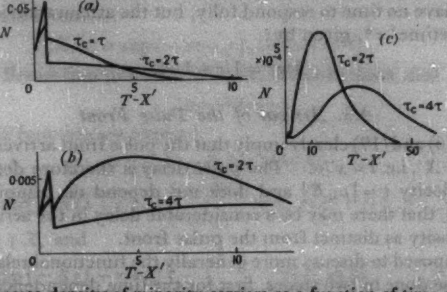


Figure 3. The excess density of minority carriers as a function of time at different points, with $T_1 = 0{\cdot}625$ and $\tau_s = \frac{1}{2}\tau$. (a) $X' = 1$, $m = 8$; (b) $X' = 2$, $m = 16$; (c) $X' = 4$, $m = 32$.

4.5. *Spatial Distribution of the Free and Trapped Carriers at a Given Time*

The spatial distribution of free and trapped carriers at a given time is of interest, especially when evaluating the photoconductive response, where the total number of added carriers present in the filament at any time is required. Curves for $N(X, T)$ and $N_t(X, T)$, with T as a parameter, have been computed for various sets of values of the parameters τ_s, τ_c and T_1. The results for two such sets, representing rather extreme conditions, are shown in figures 4 and 5. Figure 4 corresponds to 'moderate trapping', $\tau_c/\tau_s = 1$, while figure 5 gives the curves for the case of 'relatively heavy trapping', $\tau_c/\tau_s = 50$. In the former case the trapped and free densities appear to move in space with a slight shift between them. In the latter case, however, both the free and trapped charge densities are effectively 'locked' near the injection point so that the pattern changes but very little after times as long as $T=20$ and 30. The free charge density in the 'tail' is quite negligible in this case, apart from the actual pulse, compared with the trapped charge density, thus justifying the neglect of the former in computing the photo-response.

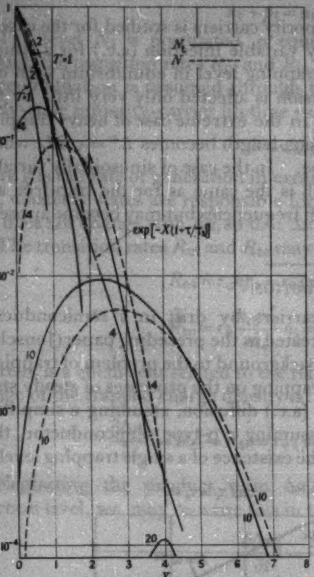


Figure 4. The excess free carrier density N and trapped carrier density N_t, as a function of distance at different times. 'Moderate trapping': $\tau/\tau_s=\tau/\tau_c=a=1$, $\tau_c/\tau_s=1$, $T_1=2$.

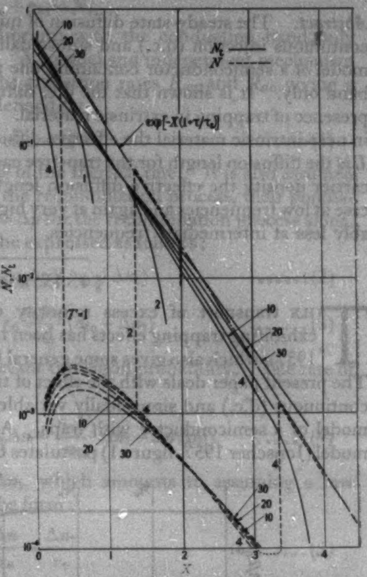


Figure 5. The excess free carrier density N and trapped carrier density N_t, as a function of distance at different times. 'Relatively heavy trapping': $\tau/\tau_s=1$, $\tau/\tau_c=a=0{\cdot}02$, $\tau_c/\tau_s=50$, $T_1=0{\cdot}625$.

One interesting feature of figure 5 is the cross-over of the curves for free carrier density. The implication is that a collector point placed to the left of the cross-over point will record an excess charge density dropping after the passage of the pulse, as indicated in figure 3 (*a*), while a collector placed to the right of the cross-over point in the same specimen will record a rising characteristic rather like those of figures 3 (*b*) and (*c*).

REFERENCES

Fan, H. Y., 1953, *Phys. Rev.*, **92**, 1424; 1954, *Physica*, **20**, 855.
Haynes, J. R., and Hornbeck, J. A., 1955, *Phys. Rev.*, **100**, 606.
Hornbeck, J. A., and Haynes, J. R., 1955, *Phys. Rev.*, **97**, 311.
Rittner, E. S., 1956, contained in *Photoconductivity* (Ed. Breckenridge, Russel and Hahn) (New York: Wiley).
Shockley, W., and Read, W. T., 1952, *Phys. Rev.*, **87**, 835.

Diffusion of Minority Carriers in the Presence of Trapping

By A. K. JONSCHER
Research Laboratories of The General Electric Company Ltd., Wembley, Middx.

MS. received 5*th November* 1956, *and in final form* 3*rd December* 1956

Abstract. The steady-state diffusion of minority carriers is studied for the case of continuous injection (d.c.) and sinusoidally variable injection (a.c.) for a simple model of a semiconductor containing one trapping level in equilibrium with one band only. It is shown that the d.c. diffusion is affected only very little by the presence of trapping in extrinsic material. In the extreme case of heavy trapping in near-intrinsic material the effective diffusion length becomes $L^* = \sqrt{2}L$, where L is the diffusion length for the trap-free case. In the case of sinusoidally variable carrier density the effective diffusion length is the same as for the trap-free a.c. case at low frequencies and again at very high frequencies but may become appreciably less at intermediate frequencies.

---

§ 1. INTRODUCTION

THE transport of excess minority carriers by drift in a semiconductor exhibiting trapping effects has been treated in the preceding paper (Jonscher 1957), which also gives some general background to the problem of trapping. The present paper deals with the effect of trapping on the processes of steady state continuous (d.c.) and sinusoidally variable (a.c.) diffusion, assuming a simplified model of a semiconductor with traps. Assuming a p-type semiconductor, this model (Jonscher 1957, figure 1) postulates the existence of a single trapping level of

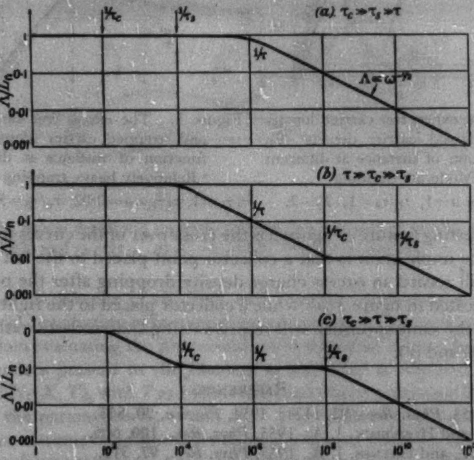


Figure 1. The dependence of the effective a.c. diffusion length Λ on frequency ω for different values of τ_c and τ_s in relation to a fixed value of lifetime τ (qualitative only).

density N_1. The concentrations of electrons and holes are as follows: $n=n_0+\Delta n$ electrons in the conduction band, $p=p_0+\Delta p$ holes in the valance band, $n_t=n_{t0}+\Delta n_t$ electrons in the trapping level, $p_t=p_{t0}-\Delta n_t$ holes in the trapping level. The subscripts 0 refer to thermal equilibrium densities and the Δ's denote added concentration in excess of thermal equilibrium. We have the obvious relation

$$n_t+p_t=n_{t0}+p_{t0}=N_1.$$

The level N_1 is postulated in equilibrium with the conduction band only through the transition rates R_{ct} and R_{tc}. A parallel and independent mechanism of recombination is assumed through the transition rates R_{cv} and R_{vc} giving a steady-state minority-carrier lifetime τ defined by

$$\Delta n_{\text{steady state}}=\tau g \qquad \ldots\ldots(1)$$

where g is the rate of generation in excess of the thermal rate. It is also assumed that, whatever states are responsible for the recombination process, their population does not change with Δn, so that $\Delta n=\Delta p$ for the recombination process alone.

The transition rates R_{ct} and R_{tc} may be expressed as follows:

$$R_{ct}=r_c n p_t=r_c(n_0+\Delta n)(p_{t0}-\Delta n_t) \qquad \ldots\ldots(2)$$

$$R_{tc}=r_c\frac{n_0 p_{t0}}{n_{t0}}n_t=r_c\frac{n_0 p_{t0}}{n_{t0}}(n_{t0}+\Delta n_t) \qquad \ldots\ldots(3)$$

where r_c is a coefficient related to the relevant transition probability. The rate of change of the trapped charge density is:

$$\frac{\partial \Delta n_t}{\partial t}=R_{ct}-R_{tc}=r_c\left(p_{t0}\Delta n-\frac{n_0 N_1}{n_{t0}}\Delta n_t-\Delta n\,\Delta n_t\right). \qquad \ldots\ldots(4)$$

Neglecting the product term $\Delta n\,\Delta n_t$, which amounts to assuming a low injection level, we may re-write this in the form:

$$\frac{\partial \Delta n_t}{\partial t}=\frac{\Delta n}{\tau_s}-\frac{\Delta n_t}{\tau_c} \qquad \ldots\ldots(5)$$

where

$$\frac{1}{\tau_s}=r_c p_{t0} \qquad \ldots\ldots(6)$$

$$\frac{1}{\tau_c}=\frac{r_c n_0 N_1}{n_{t0}}. \qquad \ldots\ldots(7)$$

The continuity equation for the free minority carriers is

$$\frac{\partial \Delta n}{\partial t}=-\frac{\Delta n}{\tau}+D_n\frac{\partial^2 \Delta n}{\partial x^2}+\mu_n E\frac{\partial \Delta n}{\partial x}-\frac{\partial \Delta n_t}{\partial t}+g \qquad \ldots\ldots(8)$$

where D_n and μ_n are the electronic diffusion constant and mobility respectively.

§ 2. STEADY STATE DIFFUSION WITH CONTINUOUS INJECTION

Consider a constant density of excess carriers (d.c. case) being maintained at the boundary of a semi-infinite semiconducting body, whether by injection from a p–n junction or generation by light. No electric field is applied externally to the material, but the field E is generally different from zero and is determined by the boundary conditions imposed on the particle currents of holes and electrons, J_p

and J_n respectively. In the case of generation by light the condition is that both currents should be equal, so that the total current should be zero. The potential difference arising in the specimen is then called the photo-diffusion voltage (Moss *et al.* 1953). In the case of an emitter junction the ratio of the particle currents is determined by the emitter efficiency γ. We propose to consider the former case, in which

$$J_p \equiv \mu_p p E - D_p \frac{\partial \Delta p}{\partial x} = -\mu_n n E - D_n \frac{\partial \Delta p}{\partial x} \equiv J_n. \qquad \ldots\ldots(9)$$

If $b = D_n/D_p$ is different from unity, as is normally the case, a field E will be established to take account of the difference of mobilities. In the presence of trapping it may be argued that we have a third class of carriers whose mobility is zero and which may, therefore, affect the diffusion process of the free carriers. It is this effect which we propose to investigate.

From equation (9) follows:

$$E = \frac{D_n}{\mu_n} \frac{\partial \Delta p/\partial x - b\, \partial \Delta n/\partial x}{p + bn}. \qquad \ldots\ldots(10)$$

The condition of strict neutrality in the steady state is, using (5),

$$\Delta p = \Delta n + \Delta n_t = \Delta n\,(1 + \tau_c/\tau_s). \qquad \ldots\ldots(11)$$

It may be assumed that the appearance of a small field E will not be accompanied by any appreciable departures from strict neutrality and therefore it is permissible to re-write equation (10) thus:

$$E = \frac{D_n}{\mu_n} \frac{(1 + \tau_c/\tau_s - b)\, \partial \Delta n/\partial x}{p_0 + bn_0 + (1 + \tau_c/\tau_s + b)\, \Delta n}. \qquad \ldots\ldots(12)$$

Substituting this in (8) we get the steady-state differential equation for minority carriers:

$$\frac{\partial^2 \Delta n}{\partial x^2} = \frac{\Delta n}{L_n^2} - \frac{(\partial \Delta n/\partial x)^2}{\Pi + B\Delta n} \qquad \ldots\ldots(13)$$

where $L_n = (D_n \tau)^{1/2}$, $\Pi = \dfrac{p_0 + bn_0}{1 + \tau_c/\tau_s - b}$, $B = \dfrac{1 + \tau_c/\tau_s + b}{1 + \tau_c/\tau_s - b}$.

In strongly extrinsic material the second term in equation (13) may be neglected and we obtain the solution

$$\Delta n = (\Delta n)_{x=0} \exp(-x/L_n) \qquad \ldots\ldots(14)$$

which is exactly the same as that applicable to diffusion in the absence of trapping.

If trapping is relatively heavy, i.e. if in the steady state $\Delta n_t/\Delta n = \tau_c/\tau_s \gg 1$ we may write $B = 1$. If, moreover, the material is near-intrinsic, the condition $\Pi \ll B\Delta n$ may be satisfied relatively easily. Under these extreme conditions, therefore, equation (13) reduces to

$$\Delta n \frac{\partial^2 \Delta n}{\partial x^2} + \left(\frac{\partial \Delta n}{\partial x}\right)^2 = \frac{(\Delta n)^2}{L_n^2} \qquad \ldots\ldots(15)$$

which may be written

$$\frac{\partial^2}{\partial x^2} (\Delta n)^2 = \left(\frac{2}{L_n^2}\right) (\Delta n)^2.$$

The solution of this is

$$\Delta n = (\Delta n)_{x=0} \exp(-x/L_n^*) \qquad \ldots\ldots(16)$$

where $L_n^* = \sqrt{2}\, L_n$ is the effective diffusion length for free minority carriers in the presence of relatively heavy trapping in near-intrinsic material.

Thus under these extreme conditions the influence of trapping on the effective diffusion length of free minority carriers is to increase it by a factor $\sqrt{2}$ compared with trap-free material.

It is reasonable to conclude, from physical considerations, that the extreme conditions postulated above, viz. heavy trapping and near-intrinsic material, might lead to the maximum disturbance of the diffusion process compared with trap-free diffusion. This suggests that in all practically important cases the effective diffusion length for free carriers lies between L and $\sqrt{2}L$, where L is the trap-free diffusion length for the carrier in question.

One is, therefore, led to the conclusion that trapping affects the d.c. diffusion process only very little or not at all. This statement has a direct bearing on the measurement of lifetime by diffusion (light-spot, line-of-light) methods in semiconductors exhibiting trapping. These methods would be expected, in principle, to yield the correct value of minority carrier lifetime, irrespective of trapping (Hogarth 1956).

§ 3. STEADY STATE DIFFUSION WITH SINUSOIDALLY VARIABLE INJECTION

In this case (a.c. diffusion) we are considering the same semiconducting material and we assume the boundary condition to be $(\Delta n)_{x=0} = A e^{i\omega t}$ i.e., a sinusoidal variation with an angular frequency ω. The solutions will be of the form:

$$\Delta n = f(x)\, e^{i\omega t}, \quad \Delta n_t = h(x)\, e^{i\omega t}. \qquad \ldots\ldots(17)$$

The differential equations determining the functions $f(x)$ and $h(x)$ are obtained from (8) and (5), thus

$$\left(i\omega + \frac{1}{\tau}\right) f(x) = D_n f''(x) - i\omega h(x) \qquad \ldots\ldots(18)$$

$$\left(i\omega + \frac{1}{\tau_c}\right) h(x) = \frac{1}{\tau_s} f(x) \qquad \ldots\ldots(19)$$

where dashes denote differentiation with respect to the argument. In these equations we have neglected the drift term from equation (8) in order to make a simple solution possible. This step appears to be justified in view of the conclusion of the foregoing section that the field does not affect the diffusion very appreciably.

Elimination of $h(x)$ from (18) and (19) leads to

$$f''(x) = f(x)/\Lambda^2 \qquad \ldots\ldots(20)$$

where

$$\frac{1}{\Lambda^2} = \frac{1}{L_n^2}\left(1 + i\omega\tau + \frac{i\omega\tau/\tau_s}{i\omega + 1/\tau_c}\right). \qquad \ldots\ldots(21)$$

We thus have the following results:

$$\Delta n = A\, e^{-x/\Delta}\, e^{i\omega t}, \quad \Delta n_t = \frac{A/\tau_s}{i\omega + 1/\tau_c}\, e^{-x/\Delta}\, e^{i\omega t}. \qquad \ldots\ldots(22)$$

It follows that the signal Δn is propagated with a diffusion length Λ which is complex and different from the steady state d.c. diffusion length L_n. There is also a phase shift between Δn and Δn_t, the free and trapped carrier densities.

Figure 1 (*a*), (*b*) and (*c*) shows qualitatively the depenednce of Λ on the angular frequency ω for three different relations between τ, τ_c and τ_s, all corresponding to $\tau_\mathrm{c}/\tau_\mathrm{s} \gg 1$, i.e. to 'relatively heavy trapping'. Case (*a*), the lifetime τ being shorter than τ_c and τ_s, is undistinguishable from the trap-free case (Shockley 1950, p. 316). Cases (*b*) and, especially, (*c*) show that the effective diffusion length for a.c. may be expected to fall at lower frequencies if the lifetime τ is longer than or intermediate between τ_c and τ_s. At very high frequencies all three cases again become undistinguishable from the trap-free behaviour.

REFERENCES

HOGARTH, C. A., 1956, *Proc. Phys. Soc.* B, **69**, 791.
JONSCHER, A. K., 1957, *Proc. Phys. Soc.* B, **70**, 223.
MOSS, T. S., PINCHERLE, L., and WOODWARD, A. M., 1953, *Proc. Phys. Soc.* B, **66**, 743.
SHOCKLEY, W., 1950, *Electrons and Holes in Semiconductors* (New York : van Nostrand).

The Electron Bombardment of Thin Barium Films [†]

By L. JACOB

George Holt Physics Laboratories, University of Liverpool

MS. received 12*th November* 1956

Abstract. Experiments are described which show that it is possible to transfer barium from an anode as a thin film to a cathode by simple electron bombardment. This effect appears to occur with the anode cold. The 'threshold' bombarding energy is about 300 volts for barium; the corresponding current density at the target is of the order of 1 mA cm^{-2}. The gradual build-up of thin film activity of a pure strontium oxide cathode was followed with time. The final result was an emitter whose performance matched that of the standard BaO/SrO cathode in every way; the thin film emitter, however, exhibits a large temperature dependence but has high tenacity.

§ 1. INTRODUCTION

THE origin of the series of experiments to be described in this paper arose from an observation of the author (1946) on a cathode-ray tube electron gun and focusing system. Here, the final anode (of the focusing lens) had been coated with a thin film of barium azide before processing, to see whether electron bombardment of the anode under normal operating conditions was capable of releasing free barium from it. After taking the tube off the pump it was found impossible to obtain the expected standard emission from the cathode. X-ray examination of the coating showed that whereas the original composition of the standard BaO/SrO mixture was approximately 50/50 Ba/Sr, the ratio now stood at 70/30; this would account for the loss of emission observed. As the cathode subtended an angle of about 1° at the final anode, it was surmised that the additional barium had come from the bombarded electrode. That positive ions rather than uncharged particles were involved, followed from the marked pitting of the coating and the comparatively large positive ion currents which flowed to the modulating electrode, when it was held at a negative potential with respect to the cathode.

It is known that the 'cardinal' points of such lens fields depend on geometry and voltage only, and are thus the same for particles of either sign and any mass. The process of transfer hence appears to involve the movement of positive ions back to the emitting source; indeed, the variation in emission of the source during the transfer can be watched as the process of building up a thin film emission on the cathode progresses.

Where the target material adhered loosely to it, a characteristic colour of the metallic vapour (e.g. green with barium) was visible when the current density on the anode was sufficiently high. In all cases where this 'instantaneous vacuum distillation' occurred, the cathode was found to be immediately sensitized, though the temperature of the target did not greatly exceed 100°c. The vapour pressure involved at this temperature could not possibly account for the formation of thin film activity on the surface of the emitter.

[†] Based on a paper read at the First European Conference on Electron Microscopy: Stockholm, September 1956.

Modulation characteristics taken during the build-up process showed distinct changes of slope with time, and it appeared possible to produce an emitter giving a current density greater than that obtained from the standard BaO/SrO coating.

§ 2. Experimental Details

It is a well known fact that space charge limited currents obtained from an oxide-coated cathode depend on the state of its surface. This is reflected in its emission properties for a constant field at its surface; the latter can be achieved in an electron gun by adjusting the accelerator potential to give a constant value for the cut-off when the modulator is biased negatively with respect to the cathode —in the present series of experiments −30 volts. The current furnished with the modulator at cathode potential then gives a direct measure of the quality of the emitter. Any sensitization of the surface produced by electron bombardment of a suitable target can thus be assessed in terms of the subsequent emission under these conditions.

The design of the experimental tube is shown in figure 1. The conventional gun consisted of an indirectly heated cathode shell C of nickel, of external diameter 1·7 mm (0·1 mm wall) on whose closed end was placed the test oxide emitter of pure strontium carbonate, a modulating electrode M of 10 mm diameter (insulated from the cathode by mica washers and spaced 0·2 mm from the cathode surface) containing a 1 mm diameter hole at its centre and an accelerating etectrode A in the form of a disc with a 2 mm hole at its centre coaxial with the cathode and modulator and spaced 1 mm from the latter. Because of a slight spread in dimensions during assembly, slightly different values of accelerator voltage were required for the 30 volt cut-off on various units, ostensibly made to the same specifications, but generally it required some three hundred volts on the accelerator to give a 30 volt cut-off for the beam. The target T consisted of a nickel disc 25 mm in diameter held on a lead which passed through a ' pinch ' at the opposite end of the tube to that of the gun; it was spaced some 50 mm from the accelerator. Several ' getter ' discs could be ' fired ' to deposit a coherent film of electropositive barium on the surface of the target. An interceptor test anode I held on a piece of soft iron was used as the collector for the electron emission from the cathode after the various stages of bombardment of the target.

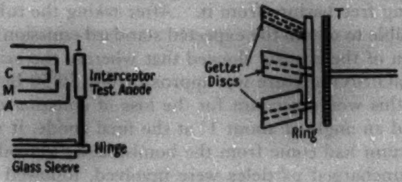


Figure 1.

The processing was carefully controlled. The cathode was heated to decompose the strontium carbonate to the oxide and space current drawn to the test anode before finally bringing down the getter and sealing off from the pump. In some instances a series of experiments lasting over several weeks was carried out while the tube was on the pump.

The following procedure was adopted in making the measurements :

The modulator was kept at cathode potential and the voltage of the accelerator A set at 100. The target potential was raised, starting at 100 volts, in 100 volt steps to 300 volts, for various intervals of time. At the end of each such interval, the interceptor anode I was put into position to receive the electron beam and the potential of both A and I kept at approximately 300 volts (for the 30 volt cut-off) : the total cathode current with the modulator at cathode potential was recorded.

For target potentials above 300 volts, the accelerator was kept permanently at 300 volts ; in almost all cases the target potential was kept above that of the accelerator to provide a field favourable for the return of positive ions to the emitter. The cleaning of the emissive surface was accomplished by moving the test target out of the path of the beam and flashing the cathode at 1150°c till the original low emission was reached.

The three objectives aimed at were (1) to look for the existence of a threshold for bombarding energy and current density respectively (2) to see if the thin film ' build-up ' was sensitive to temperature changes, (3) to investigate such properties as stability and tenacity of the film.

§ 3. BARIUM ON STRONTIUM OXIDE

It is a known fact that with standard BaO/SrO emitters of the dimensions used here, an emission of about 500 μA is obtained (at normal operating temperatures) for a cut-off voltage of 30 volts; small variations in temperature do not appreciably affect this value. With the pure SrO emitter it was comparatively easy to follow any growth of activity consequent on the transfer of barium from the target to the emitter. Figure 2(*a*) shows the build-up with time of a barium film from a ' thick ' target during the early stages of the bombardment, beginning at 350 volts; the activation is complete at 800 volts at which the emission far exceeds the standard value and can only be reduced by lowering the temperature. When this is done, it is possible to build up a further film at reduced potential as shown in figure 2(*b*) for the 450 volt case, but this takes much longer—in the present instance well over an hour; an immediate activation occurs, however, if the voltage is increased to 600 v after 110 minutes at 450 volts. At this stage further reduction in temperature enables additional activation to occur, but there is a definite limit to the reduction in temperature. This can be seen from figure 2 (*c*), where bombardment up to 1000 volts (which undoubtedly releases barium from the target) produces no further effect. This indicates that temperature is an important factor controlling the enhanced emissive property. The latter can be destroyed by flashing at 1050°c during which the surface barium evaporates leaving the pure SrO base; the cycle can be repeated by subsequent bombardment at 600 volts as shown in the same figure.

This cycle of alternation as between poor and good quality emission can be induced by repetition of the steps described above. It results in an emitter which appears to be somewhat more sensitive to temperature changes than that shown by the standard oxide coating. In spite of this, the film when once ' built up ' seems fairly permanent, since flashing at 1050°c for over three hours hardly affects the emission; in figure 2 (*c*) the time scale has been altered to show this. Such behaviour is unexpectedly good, for x-ray analysis shows that standard cathodes lose some 90% of their barium after one hour's flashing at this temperature.

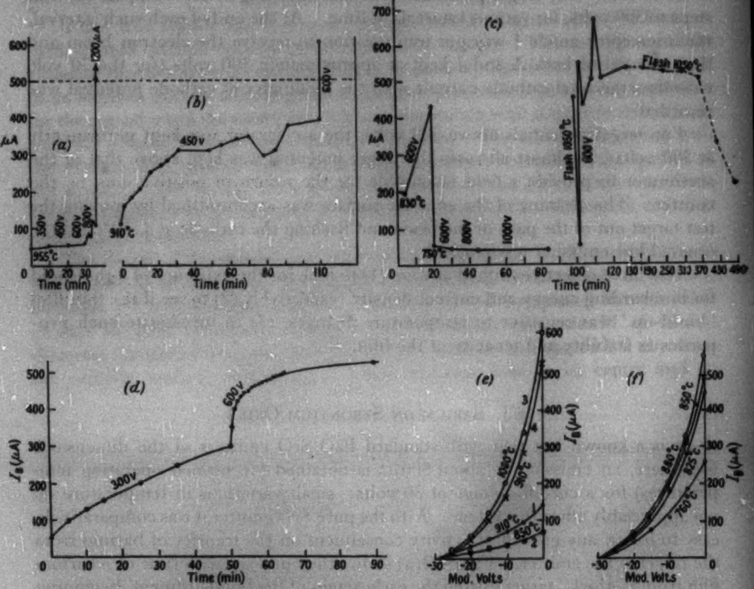


Figure 2. (*a*), (*b*), Barium on SrO: early stages of activation; (*c*), barium on SrO; (*d*), barium getter on target; (*e*), (*f*), barium on SrO: modulation characteristics; (*e*) partly and fully formed film, (*f*) thick film.

The experiments with very thin films of dispersed barium ' getter ' showed that it was possible to induce a slow partial activation with the target at 300 volts over a period of 50 minutes; at 600 volts, the activation was complete after a further 15 minutes (figure 2 (*d*)). A complication arises here in assessing the time taken to form the thin barium film on the emitter, since it is likely that the rate of activation depends on the bombarding current density and this increases with time as the activation proceeds. It can, however, be stated, that in the case of barium the threshold bombarding energy does not appear to be below 300 volts for the initial bombarding current densities, of the order of 1 mA cm^{-2} used here.

The build-up of the film was followed through its various stages by plotting the modulation characteristics of the gun. These are shown in figure 2 (*e*), curve 1, for the partly formed film, induced by bombardment of the target at

350 volts. In this condition the emitter is exceedingly sensitive to temperature changes; for example, reducing its temperature to 850°c gave curve (2), while increasing it to 1060°c produced curve (3). After full activation at this bombarding energy, curve (4) was obtained, that is, a curve similar to that given by a standard BaO/SrO cathode. This excessive sensitivity to temperature change is specific to this type of built-up film emission, and is largely different from the behaviour of normal BaO/SrO emitters. If one proceeds to thicken the film further by raising the bombarding energy to, say, 800 volts, then it is possible to achieve the values recorded for the standard characteristic, but at a temperature appreciably less than that required with the standard mixture of oxides. This is shown in figure 2 (*f*).

ACKNOWLEDGMENT

The author wishes to thank Professor H. W. B. Skinner for facilities for carrying out some of the measurements.

REFERENCE

JACOB, L., 1946, *Nature, Lond.*, **157**, 586.

RESEARCH NOTES

The Galvanomagnetic and Thermomagnetic Effects in Semiconductors

By R. MANSFIELD
Bedford College, London

MS. received 12*th November* 1956

§ 1. Introduction

The object of this note is to calculate values of the galvanomagnetic and thermomagnetic coefficients for a semiconductor in which the charge carriers are scattered by ionized impurity centres. In order to calculate these coefficients it is necessary to evaluate various integrals which are dependent on the relaxation time. If the variation of the relaxation time with energy of the charge carriers obeys a simple power law, then these integrals can be evaluated exactly. For scattering of the charge carriers by thermal vibrations of the lattice, the relaxation time varies as the reciprocal of the square root of the energy, and values of the various galvanomagnetic and thermomagnetic effects have been determined for this case. When impurity scattering of the charge carriers is considered, the relation between relaxation time and energy is more complicated and it is no longer possible to derive explicit values of the integrals.

An approximate method for calculating the integrals has been used in numerous investigations of the transport phenomena of semiconductors by assuming a simple power-law function of relaxation time with energy of the form $\tau = bE^{3/2}$. Since many of the properties considered can be expressed as ratios of the integrals, results have been obtained in which the constant b cancels, giving rise to considerable simplification of the formulae. However, in a recent latter (Mansfield 1956) the use of this approximation for calculating the Hall coefficient was considered. It was shown that, although the quantity b is a slowly varying function of energy and in certain cases may be considered a constant, its value depends on the integrands of the integrals whose values are required. Consequently, the factor b has to be retained in the expression for the Hall coefficient. The error introduced by assuming b is constant was also investigated and it was pointed out that for calculating the other properties of semiconductors this assumption might be unsatisfactory.

§ 2. Theory

Formulae for the galvanomagnetic and thermomagnetic effects for semiconductors can be derived from the equations given by Wilson (1953) for the electric and thermal currents along a conductor in which there is a longitudinal thermal gradient and a transverse magnetic field, provided extra terms are introduced to allow for the lattice contribution to the thermal currents.

If it is assumed that the semiconductor is isotropic, having charge carriers of one sign which are non-degenerate and are scattered by ionized impurity centres, the galvanomagnetic and thermomagnetic effects in a small magnetic field are given by the following expressions ($eH\tau \ll mc$, gaussian system of units):

Hall coefficient

$$R = -\frac{3\sqrt{\pi}\,A_2^{9/2}}{4n\mathbf{e}\mathbf{c}(A_1^3)^2} = -\frac{r}{n\mathbf{e}\mathbf{c}};$$

Nernst coefficient

$$Q = \frac{3\sqrt{\pi}\sigma\mathbf{k}}{4n\mathbf{e}^2\mathbf{c}}\left(\frac{A_1^3 A_1^{11/2} - A_1^4 A_2^{9/2}}{(A_1^3)^3}\right) = \frac{\sigma\mathbf{k}}{n\mathbf{e}^2\mathbf{c}}q;$$

Ettingshausen coefficient

$$P = \frac{QT}{K};$$

Righi–Leduc coefficient

$$S = \frac{3\sqrt{\pi}\mathbf{k}^2 T\sigma^2}{4n\mathbf{e}^3\mathbf{c}K}\,\frac{(2A_1^3 A_1^4 A_2^{11/2} - (A_1^3)^2 A_2^{13/2} - (A_1^4)^2 A_2^{9/2})}{(A_1^3)^4}$$

$$= -\frac{\mathbf{k}^2 T\sigma^2}{n\mathbf{e}^3\mathbf{c}K}s;$$

Change in electrical conductivity in a magnetic field is

$$\frac{\sigma_0 - \sigma}{\sigma_0} = \left(\frac{H\sigma_0}{n\mathbf{e}\mathbf{c}}\right)^2 \frac{9\pi}{16}\left(\frac{A_3^6}{(A_1^3)^3} - \left[\frac{A_2^{9/2}}{(A_1^3)^2}\right]^2\right)$$

$$= \left(\frac{H\sigma_0}{n\mathbf{e}\mathbf{c}}\right)^2 t;$$

Thermoelectric power ($H=0$)

$$\frac{d\Theta}{dT} = \frac{\mathbf{k}}{\mathbf{e}}(\eta^* - \beta), \qquad \beta = \frac{A_1^4}{A_1^3};$$

Thermal conductivity K ($H=0$)

$$K = K_{\mathrm{L}} + \lambda\left(\frac{\mathbf{k}}{\mathbf{e}}\right)^2 T\sigma, \qquad \lambda = \frac{A_1^5 A_1^3 - (A_1^4)^2}{(A_1^3)^2};$$

where

$$A_i^j = \int_0^\infty \frac{\eta^j\,\mathrm{e}^{-\eta}\,d\eta}{(f(a\eta))^i},$$

and the following formula has been used for the relaxation time:

$$\tau = \frac{\sqrt{(2\mathbf{m})}\,\epsilon^2(\eta\mathbf{k}T)^{3/2}}{\pi n\mathbf{e}^4 f(a\eta)},$$

$$f(a\eta) = \log(1 + a\eta) - \frac{a\eta}{1 + a\eta}, \qquad a = \frac{2\mathbf{m}\epsilon(\mathbf{k}T)^2}{\hbar^2\pi\,\mathbf{e}^2 n}.$$

n, η and η^* are the concentration, the reduced energy and the reduced chemical potential, respectively, of the charge carriers; ϵ is the dielectric constant, σ the electrical conductivity and K_{L} the lattice contribution to the thermal conductivity.

The above relations have been calculated for the case of conduction by electrons and the sign convention used is that given by Gerlach (1928). The signs of R and S are reversed when conduction is by positive holes. It should be noted that the sign of the Nernst effect is positive for impurity scattering, whereas for lattice scattering of the charge carriers it is negative.

Values of the integrals A_i^j have been calculated for various values of a, firstly by numerical integration and secondly by assuming that $f(a\eta)$ can be replaced by a constant using the value of η which makes the integrand a maximum, thus

$$A_i^j = \frac{j\,!}{[f(aj)]^i}.$$

Numerical Method

a	0·1	0·3	1	5	10	100
A_1^3	256·4	43·02	10·02	3·236	2·388	1·248
A_1^4	572·3	109·7	30·41	11·34	8·636	4·733
A_1^5	1905	405·0	126·6	51·70	40·16	22·73
$A_2^{9/2}$	$5{\cdot}298\times10^4$	1417	84·45	11·07	6·470	1·966
$A_2^{11/2}$	$1{\cdot}009\times10^5$	3673	313·9	51·15	30·98	10·05
$A_2^{13/2}$	$3{\cdot}137\times10^5$	$1{\cdot}442\times10^4$	1551	290·9	181·3	61·52
A_3^6	$12{\cdot}16\times10^6$	$4{\cdot}772\times10^4$	809·3	50·69	24·02	4·517
r	1·071	1·017	1·118	1·405	1·508	1·677
q	−0·351	0·044	0·763	1·57	1·77	2·22
s	2·56	3·52	5·61	8·68	9·73	11·6
t	0·127	0·025	0·172	0·670	0·843	1·299
β	2·23	2·55	3·03	3·50	3·62	3·79
λ	2·45	2·91	3·43	3·70	3·74	3·83

Approximate Method

a	0·1	0·3	1	5	10	100
A_1^3	189·9	35·67	9·429	3·270	2·430	1·274
A_1^4	472·8	98·76	29·65	11·47	8·762	4·803
A_1^5	1663	379·4	125·2	52·24	40·64	23·00
$A_2^{9/2}$	$1{\cdot}396\times10^4$	668·2	66·58	10·81	6·439	2·002
$A_2^{11/2}$	$4{\cdot}136\times10^4$	2324	273·4	50·62	31·05	10·19
$A_2^{13/2}$	$1{\cdot}640\times10^5$	$1{\cdot}056\times10^4$	1420	289·4	182·2	62·27
A_3^6	$8{\cdot}397\times10^5$	$1{\cdot}244\times10^4$	557·7	48·00	23·54	4·571
r	0·514	0·698	0·995	1·344	1·447	1·639
q	0·243	0·495	0·958	1·58	1·76	2·16
s	1·65	2·95	5·36	8·36	9·46	11·4
t	−0·0475	−0·0025	0·1857	0·620	0·797	1·219
β	2·49	2·77	3·14	3·51	3·61	3·77
λ	2·56	2·97	3·39	3·67	3·72	3·84

The table gives the values of these integrals together with the values of the galvanomagnetic and thermomagnetic coefficients. Figure 1 compares the results obtained by these two methods, except for the Hall coefficient which has been considered previously.

Because of the reversal in sign of the Nernst coefficient, the transition from impurity to lattice scattering was examined for this effect. When both lattice and impurity scattering occurs, the relaxation time can be written as

$$\tau = \frac{C_{\mathrm{L}}\,\eta^{3/2}}{\eta^2+\alpha}, \qquad \text{where} \qquad \alpha = \frac{C_{\mathrm{L}} f(a\eta)}{C_{\mathrm{I}}}$$

and the separate relaxation times are given by

$$\tau_{\mathrm{L}} = C_{\mathrm{L}}\,\eta^{-1/2}, \qquad \tau_{\mathrm{I}} = C_{\mathrm{I}}\,\eta^{3/2}/f(a\eta).$$

The integrals determining the Nernst coefficient for this case are of the form

$$B_i^j = \int_0^\infty \frac{\eta^j\,\mathrm{e}^{-\eta}\,d\eta}{(\alpha+\eta^2)^i}, \qquad q = \frac{B_1^3\,B_2^{11/2} - B_1^4\,B_2^{9/2}}{(B_1^3)^3}.$$

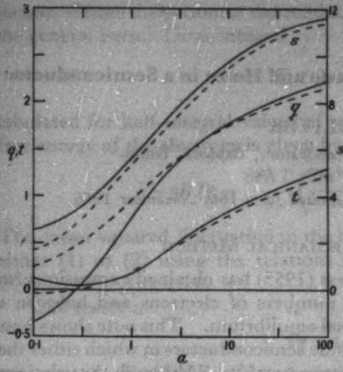


Figure 1.

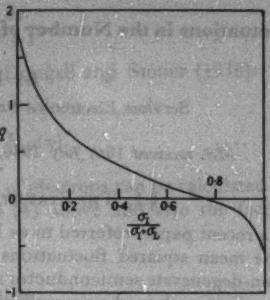


Figure 2. Nernst effect. Transition from impurity scattering to lattice scattering.

Since the values of q obtained by treating $f(a\eta)$ as a constant agreed with that obtained by numerical integration when $a = 10$, this approximation was used. The appropriate values of $\alpha_{j,i}$ for the various integrals required were obtained by deducing the value of $\eta = \eta_{j,i}$ which make the integrands a maximum. $\eta_{j,i}$ is determined by the solution of the equation

$$\eta_{j,i}^3 - (j-2i)\eta_{j,i}^2 + \alpha_{j,i}\eta_{j,i} - j\alpha_{j,i} = 0.$$

For the case $j = 3$, $i = 1$, the integral determines the electrical conductivity and it is possible to introduce the parameter σ_L/σ_I where $\alpha_{3,1} = 6\sigma_L/\sigma_I$;

then
$$\alpha_{j,i} = \alpha_{3,1}\frac{f(a\eta_{j,i})}{f(a\eta_{3,1})}.$$

Various values of σ_L/σ_I were chosen and, by successive approximations, the appropriate values of $\alpha_{j,i}$ were obtained for the particular case of $a = 10$. That is, putting $\alpha_{j,i} = \alpha_{3,1}$, $\eta_{j,i}$ was obtained, which was used to recalculate $\alpha_{j,i}$ and hence obtain a better value of $\eta_{j,i}$. The integrals B_i^j were then obtained from tables listed by Dingle and Roy, and q was calculated. Figure 2 illustrates the variations of q with $\sigma_I/(\sigma_I + \sigma_L)$. The author is indebted to Dr. Donovan for drawing his attention to these tables and making a copy of them available.

REFERENCES

GERLACH, W., 1928, *Handbuch der Physik*, Volume 13 (Berlin: Springer), p. 228.
MANSFIELD, R., 1956, *Proc. Phys. Soc.* B, **69**, 862.
WILSON, A. H., 1953, *The Theory of Metals* (Cambridge: University Press), p. 218.

Fluctuations in the Number of Electrons and Holes in a Semiconductor

By D. J. OLIVER
Services Electronics Research Laboratory, Baldock, Herts.

MS. received 18*th July* 1956, *and in revised form* 16*th November* 1956

§ 1. THE STATISTICAL MECHANICAL METHOD

In a recent paper, referred to as I, Burgess (1955) has obtained expressions for the mean squared fluctuations in the numbers of electrons and holes in a non-degenerate semiconductor in thermal equilibrium. This note shows how the previous work can be generalized to include semiconductors in which either the electrons or holes (or both) form a degenerate assembly. Although the relations obtained have general validity this work was prompted by an investigation into noise in indium antimonide. For this material the onset of degeneracy occurs for much lower electron concentrations that for other semiconductors (e.g. germanium) so that calculations based upon classical statistics must be used with caution. The problem has been discussed by the methods of statistical mechanics.

In the first approach, the mean squared fluctuation in a variable x of an assembly is given by the equation

$$\overline{(x-x_0)^2}=\boldsymbol{k}T/(\partial^2F/\partial x_0^2) \qquad \ldots\ldots(1)$$

where F is the free energy of the assembly and x_0 is the equilibrium value of x. This relation was first obtained by Einstein (1910) and a complete derivation of it on the basis of a description of the assembly by a canonical ensemble has been given by Tolman (1938). It is clear from this description (see also ter Haar 1954) that the Einstein method can be validly used with both classical and quantum statistics in contradiction to the statement of Burgess (1955).

Alternatively, a discussion of the fluctuations in the number of particles can be appropriately carried out in terms of a grand canonical ensemble leading to a second formula for the required fluctuation

$$\overline{(n-n_0)^2}=\boldsymbol{k}T/(\partial\zeta/\partial n_0) \qquad \ldots\ldots(2)$$

where ζ is the chemical potential (see Tolman 1938, p. 643). The equations (1) and (2) must clearly lead to the same results and this is apparent from the fact that, for an assembly at constant temperature, $\partial^2F/\partial n_0{}^2=\partial\zeta/\partial n_0$.

§ 2. GENERAL RESULTS

Consider an assembly of electrons occupying a standard parabolic band of energy levels. The density of electrons n is related to the Fermi energy by the equation

$$n=4\pi(2m^*\boldsymbol{k}T/\boldsymbol{h}^2)^{3/2}F_{1/2}(\eta) \qquad \ldots\ldots(3)$$

where m^* is the effective electron mass, $\eta\,(=\zeta/\boldsymbol{k}T)$ is the reduced Fermi energy measured from the bottom of the conduction band and $F_{1/2}(\eta)$ is a particular case of the general Fermi–Dirac integral

$$F_n(\eta)=\int_0^\infty \frac{x^n\,dx}{e^{x-\eta}+1}$$

tabulated for half-integral values of n by McDougall and Stoner (1938). The free energy of the electrons is given by

$$F=nkT\eta-\frac{8\pi kT}{3}\left(\frac{2m^*kT}{h^2}\right)^{3/2}F_{3/2}(\eta). \qquad \ldots\ldots(4)$$

The mean squared fluctuation in the number of electrons can be calculated from either (1) or (2) using the relations (3) and (4) above and also the fact that $F_n'(\eta)=nF_{n-1}(\eta), n>0$.

The result is

$$\overline{(n-n_0)^2}=\alpha(\eta)\,n_0 \qquad \ldots\ldots(5)$$

where $\alpha(\eta)=F_{1/2}'(\eta)/F_{1/2}(\eta)$. A plot of α as a function of η is given in figure 1. In the classical limit $\alpha(\eta)\rightarrow 1$ whilst for complete degeneracy $\alpha(\eta)\rightarrow 3/2\eta$. The fluctuations in the number of holes can be obtained in a similar way leading to the formula

$$\overline{(p-p_0)^2}=\alpha(-\eta-E_G/kT)\,p_0 \qquad \ldots\ldots(6)$$

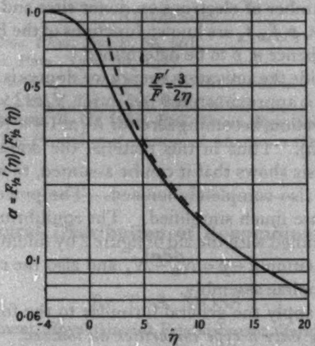


Figure 1. The degeneracy factor α as a function of the reduced Fermi energy η $(=\zeta/kT)$. The broken line indicates the values of α for complete degeneracy.

where E_G is the energy gap between the top of the valence band and the bottom of the conduction band. Putting $n-n_0=\Delta n$ and $p-p_0=\Delta p$ the equations obtained in I for $\overline{(\Delta n)^2}$ and $\overline{(\Delta p)^2}$ can be used in the discussion of degenerate samples with the substitution of $\alpha(\eta)\,n_0$ for n_0 and $\alpha(-\eta-E_G/kT)\,p_0$ for p_0 whenever these occur.

§ 3. APPLICATION TO INDIUM ANTIMONIDE

The relations between the reduced chemical potentials (η_n, η_p) and the respective densities n, p are illustrated in figure 2 for InSb at room temperature.

For a given concentration of donor impurities N_D, and acceptor impurities N_A, the difference between n and p is given quite generally by the relation

$$n - p = N_D(1 - f_D) - N_A f_A$$

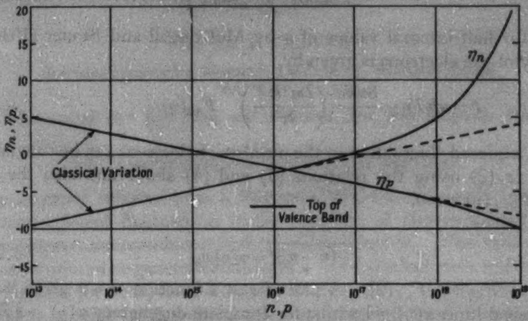


Figure 2. The reduced chemical potentials for electrons and holes, as a function of n and p respectively in InSb ($m_n^* = 0{\cdot}03m$, $m_p^* = 0{\cdot}2m$, $E_G = 0{\cdot}18$ ev, $T = 300°\text{K}$). The zero of energy is taken at the bottom of the conduction band and the top of valence band is at $-7{\cdot}2$. The dotted line indicates the classical variation $n = \text{const. } e^{\eta}$.

where $N_D f_D$ is the number of electrons on donor sites and $N_A f_A$ the number on acceptor sites. Since n, p, f_D, f_A are known functions of the Fermi level η the above relation enables η and hence n, p to be determined.

In indium antimonide the activation energy for donors is zero (Hrostowski *et al.* 1955) and for acceptors is approximately 0·005 ev for $N_A \simeq 2 \times 10^{16}$ cm^{-3}, decreasing with impurity concentration, becoming zero for $N_A \simeq 1{\cdot}4 \times 10^{17}$ cm^{-3} (Broom 1956, private communication). Thus in this material the donors are always ionized whilst a rigorous analysis shows that it can be assumed, to within an error of 5%, that the acceptors are also completely ionized. The procedure for locating the Fermi level is in this case much simplified. The equilibrium density of electrons and holes can be determined with the aid of figure 2 by finding those values of n and p which satisfy the equation $n - p = N_D - N_A$ and also the relation $\eta_n = \eta_p = \eta$ the Fermi level for the complete assembly.

It is instructive to apply the general formulae to the following special cases: (i) *A sample containing only n-type impurities all ionized.* In this case

$$\alpha(-\eta - E_G/kT) = 1$$

so that Burgess' equation (27), $\overline{(\Delta n)^2} = \overline{(\Delta p)^2} = n_0 p_0/(n_0 + p_0)$ becomes modified to read

$$\overline{(\Delta n)^2} = \overline{(\Delta p)^2} = \alpha(\eta)\, n_0 p_0 / \{\alpha(\eta)\, n_0 + p_0\}. \qquad \ldots\ldots(7)$$

For indium antimonide the differences between this expression and the classical one are always small amounting to about 5% at the most. The reason for this is that, although for increasing n_0 and η, $\alpha(\eta)$ decreases monotonically from the classical value, these changes are accompanied by a reduction in p_0 and when $\alpha(\eta)\, n_0 \gg p_0$, $\overline{(\Delta n)^2} = \overline{(\Delta p)^2} = p_0$ which is the same as the classical result. For a material with a smaller effective electron mass or a smaller energy gap the changes introduced by the more general treatment would be greater.

(ii) *A sample containing both donor and acceptor impurities.* As explained in the introduction to this section all the donors will be ionized and it can be assumed with a small error that the acceptors will also be ionized. The general expression for the fluctuations is

$$\overline{(\Delta n)^2}=\overline{(\Delta p)^2}=\frac{n_0 p_0 \alpha(-\eta-E_G/\boldsymbol{k}T)}{n_0+\alpha(-\eta-E_G/\boldsymbol{k}T)p_0}. \qquad \ldots\ldots(8)$$

Since, however, $\alpha(-\eta-E_G/\boldsymbol{k}T)p_0 \gg n_0$, $\overline{(\Delta n)^2}=\overline{(\Delta p)^2}=n_0$.

For a material with a larger acceptor activation energy than indium antimonide partial ionization would have to be considered, and Burgess' equation (26) applied to holes and modified for the case of degeneracy is

$$\overline{(\Delta p)^2}/\alpha(-\eta-E_G/\boldsymbol{k}T)p_0=(1-\gamma)/(2-\gamma) \qquad \ldots\ldots(9)$$

where γ is the fraction of the acceptors which are ionized.

It can be concluded that the result of applying the general equations (5) and (6) to indium antimonide is to produce only small changes (~5%) in estimates of $\overline{(\Delta n)^2}$ and $\overline{(\Delta p)^2}$ in n-type samples and to produce no change in corresponding calculations for p-type samples.

ACKNOWLEDGMENTS

The author is indebted to Dr. R. Barrie for helpful discussions and to the Admiralty for permission to publish this note.

REFERENCES

BURGESS, R. E., 1955, *Proc. Phys. Soc.* B, **68**, 661.
EINSTEIN, A., 1910, *Ann. Phys., Lpz.*, **33**, 1275.
TER HAAR, D., 1954, *Elements of Statistical Mechanics* (London : Constable), p. 155.
HROSTOWSKI, H. J., *et al.*, 1955, *Phys. Rev.*, **100**, 1672.
MCDOUGALL, J., and STONER, E. C., 1938, *Phil. Trans. Roy. Soc.*, **237**, 67.
TOLMAN, R. C., 1938, *The Principles of Statistical Mechanics* (Oxford: University Press), p. 636.

Theory of the Spectral Distribution of Recombination Radiation from InSb

BY T. S. MOSS

Radio Department, Royal Aircraft Establishment, Farnborough, Hants.

MS. received 2nd August 1956, *and in revised form 20th November* 1956

ALTHOUGH there have been several publications on the spectral distribution of radiation which is generated by recombination in semiconductors, little attention has yet been paid to the distribution which will actually be observed experimentally.

We will treat the case of a thin slab where the excess carriers are generated optically by highly absorbed short wavelength radiation, so that all the excess carriers can be considered to be generated at the illuminated surface, and the recombination radiation is observed from the dark surface. The quantum efficiency is taken as unity. As the absorption coefficient exceeds $2\times10^4\,\mathrm{cm}^{-1}$ for wavelengths less than $1{\cdot}6\,\mu$, the radiation from a tungsten lamp will be substantially absorbed in less than $1\,\mu$ depth. The equilibrium rate of photon generation per unit wavelength interval per unit volume of the material will be

$$r = 4n^2KQ \qquad \ldots\ldots(1).$$

and the integrated rate $R = \int_0^\infty 4n^2KQ\,d\lambda$ per unit volume (cf. Landsberg and Moss 1956), where K is the absorption coefficient for the *photoelectric* part of the absorption, Q the tabulated Planck function for radiation *in vacuo*, so that $Q = 2\pi\mathbf{c}\lambda^{-4}(\exp h\nu/\mathbf{k}T - 1) - 1$, and the refractive index n is assumed to be constant over the region around the absorption edge which is the only part of the spectrum where r is significant. Recent measurements in this laboratory show that n has the value 4·0 and that the dispersion is indeed small in this wavelength region. The excess generation rate over equilibrium due to small, equal increases δn in both hole and electron concentrations in an intrinsic specimen of electron density n_i will be

$$\delta r = (2r/n_i)\delta n. \qquad \ldots\ldots(2)$$

Correspondingly, the total excess generation rate per unit waveband will be $\delta R = (2r/n_i)\delta N$, where δN is the total increase in the number of electrons (cf. Van Roosbroeck and Shockley 1954).

This radiation will be generated in a solid angle 4π, but, due to the high refractive index, most of the radiation reaching the dark surface will be totally reflected. By integration of the Fresnel reflection coefficients over all angles of incidence, the fraction of the total generated radiation which is transmitted is found to be exactly 1%. During the experiment of Moss, Hawkins and Smith (1957), the whole of this 1% could not be collected. Their optical system collected all radiation emitted in a cone of 65°, corresponding to radiation generated in a solid angle of 0·074 steradians. By again integrating the Fresnel coefficients, the average transmission for angles of incidence within this solid angle is found to be 67%. Hence the net fraction of the total generated radiation collected was $0{\cdot}074 \times 0{\cdot}67/4\pi = 0{\cdot}004$.

All the transmitted fraction will be incident on the radiating surface at near-normal angles, and will thus have travelled a distance equal to the specimen thickness t and suffered an attenuation $\exp(-K't)$, where K' is the *total* absorption coefficient. The directly emitted radiation per unit waveband will thus be $0{\cdot}01\,\delta R\exp(-K't)$, and, similarly, the collected radiation will be

$$0{\cdot}004\,\delta R\exp(-K't).$$

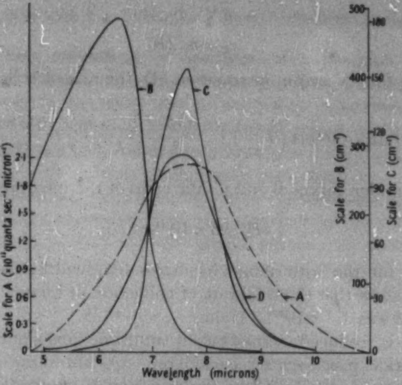
Now, provided the specimen is not a perfectly plane parallel slab, the 99% reflected radiation will travel to the illuminated surface, and so on for multiple reflections. The total collected radiation per unit waveband will, therefore, be

$$\begin{aligned}\delta R_c &= 0{\cdot}004\{\delta R\exp(-K't) + (0{\cdot}99)^2\,\delta R\exp(-3K't) \\ &\quad + (0{\cdot}99)^4\,\delta R\exp(-5K't)\} \qquad \ldots\ldots(3) \\ &= 0{\cdot}004\,\delta R\exp(-K't)\{1 - 0{\cdot}99\exp(-2K't)\}^{-1} \\ &= 0{\cdot}004\,\delta R/2\sinh K't, \text{ to a good approximation.}\end{aligned}$$

Hence $\delta R_c/A = KQ/2Q_{\max}\sinh K't, \qquad \ldots\ldots(4)$

where $Q_{\max}$ is the peak value of Q, and $A = 0{\cdot}032n^2Q_{\max}\delta N/n_i$ is independent of wavelength.

To evaluate this expression data given by Moss (1954) have been used. In this work it was found that the absorption curves were straight lines if plotted as $\log K$ against photon energy, and these straight lines have been extrapolated to find K at long wavelengths. For wavelengths greater than $8\,\mu$, K' is taken as constant at $16\,\text{cm}^{-1}$. The relevant thickness $t = 20\,\mu$.



Measured and expected recombination radiation distribution from InSb. A, measured emission curve; B, generation curve $KQ/Q_{\max}$; C, ideal emission curve $KQ/2Q_{\max} \sinh K't$; D, expected emission curve.

The function $\delta R_c/A$ which is plotted in curve C of the figure shows the radiation which should be measured experimentally with a narrow-band receiver. Curve B shows $KQ/Q_{\max}$, which is proportional to the generated radiation. It is the amount of radiation which would be collected if absorption and multiple reflection effects were ignored and may be compared directly with curve C, with allowance for the different scales used. It will be seen that the wavelength of the maximum has been shifted considerably by allowing for absorption and multiple reflections, and the peak of curve C now coincides with that of the measured curve A obtained by Moss, Hawkins and Smith (1957). Their experimental curve was broadened by the low resolution which was used. The actual passband at each wavelength was that given by a Leiss double monochromator with the slits at maximum separation, i.e. 1·5 mm. Assuming the passband to be of triangular shape, with width determined by the relevant dispersion data, we can derive from curve C the shape of the broadened curve D which should have been observed experimentally. The scale of curve D has been adjusted to make the maximum approximately the same as for curve A. It will be seen that the correspondence with the measured results is good, the worst signal discrepancy being approximately equal to the noise which was present in the system (i.e. 1/5 of maximum signal).

It should be noted that, if the specimen configurations used are such that multiple internal reflections are unimportant (e.g. the Weierstrass sphere used by Aigrain (1954)), then the emission curve will be of the simple form

$$QK \exp(-K't).$$

The estimation of the percentage of recombinations which are radiative, which equals the ratio τ_B/τ_R of the bulk lifetime to the radiative lifetime, is as follows:

If the incident illumination is such that I photons sec^{-1} are absorbed, then the assumption of unit quantum efficiency gives

$$\delta N = I\tau_B.$$

Also

$$\tau_R = n_i/2R.$$

Substituting for δN and n_i in equation (4), the rate of collection of quanta is found to be

$$\int \delta R_c = 0{\cdot}004 \left(I \frac{\tau_B}{\tau_R}\right) \int_0^\infty \frac{(QK/Q_{\max})\, d\lambda}{2 \sinh K't} \bigg/ \int_0^\infty \frac{QK d\lambda}{Q_{\max}}. \qquad \ldots\ldots(5)$$

From the areas under curves B and C, the ratio of the integrals is 0·2; therefore,

$$\int \delta R_c = 8 \times 10^{-4} I \tau_B/\tau_R. \qquad \ldots\ldots(6)$$

With allowance for the wide monochromator bandwidths used, the measured data of curve A show that the total rate of collection of quanta was $2{\cdot}6 \times 10^{13}$ per second,† when I was $1{\cdot}3 \times 10^{17}$. Hence $\tau_B/\tau_R = 0{\cdot}25$.

This figure is reduced somewhat if multiple absorption-recombination-emission-absorption processes are considered. As the ratio of the integrals in equation (4) shows that only 0·2 of the radiation is emitted, 0·8 must be self-absorbed. Most of this is absorbed photoelectrically, producing more photo-electrons and photo-holes with unit quantum efficiency. Of this, $0{\cdot}8\tau_B/\tau_R$ will reradiate, and so on. Ignoring the fact that some absorption will not be photoelectric (as this would only make a second-order correction), the total radiation will be:

$$0{\cdot}004\, I \frac{\tau_B}{\tau_R} + 0{\cdot}8 I \left(\frac{\tau_B}{\tau_R}\right)^2 + (0{\cdot}8)^2 I \left(\frac{\tau_B}{\tau_R}\right)^3 + \ldots.$$

$$= 0{\cdot}004 I \left(\frac{\tau_B}{\tau_R}\right) \left(1 - 0{\cdot}8 \frac{\tau_B}{\tau_R}\right)^{-1}. \qquad \ldots\ldots(7)$$

Substituting this value in lieu of equation (5) we obtain $\tau_B/\tau_R = 0{\cdot}21$.

We thus conclude that the above theoretical treatment—which, it may be noted, involves no arbitrary constants—gives a satisfactory explanation of the observed spectral distribution of recombination radiation from InSb. The analysis indicates that the percentage of total recombinations which are radiative is of the order of 20%.

REFERENCES

AIGRAIN, P., 1954, *Atlantic City Conference on Photoconductivity* (New York: Wiley).
LANDSBERG, P. T., and MOSS, T. S., 1956, *Proc. Phys. Soc.* B, **69**, 661.
MOSS, T. S., 1954, *Proc. Phys. Soc.* B, **67**, 775.
MOSS, T. S., and HAWKINS, T., 1956, *Phys. Rev.*, **101**, 1609.
MOSS, T. S., HAWKINS, T., and SMITH, S. D., 1957, *Report of the Semiconductor Conference* (London: Physical Society).
VAN ROOSBROECK, W., and SHOCKLEY, W., 1954, *Phys. Rev.*, **94**, 1558.

---

† With no allowance for the bandwidth correction, this figure was previously given by Moss, Hawkins and Smith (1957) and Moss and Hawkins (1956) as 6×10^{13}.

Abnormal Noise in Junction Transistors during Secondary Ionization

By R. B. JACKSON and A. K. WALTON
The Royal Aircraft Establishment, Farnborough, Hants.

MS. received 15*th June* 1956, *and in revised form* 20*th October* 1956

Under certain conditions of operation of a junction transistor when arranged in the grounded-emitter configuration, the amplification of the base current is increased due to secondary ionization. The primary current carriers are accelerated by the electric field which exists in the depletion layer at the collector junction, and acquire sufficient energy to raise some of the bound electrons from the valence to the conduction band.

While taking static characteristics of three commercial types of transistor under such conditions, excessive noise was observed at the collector junctions. The noise was of the form shown in the typical selection of photographs of figure 1 (Plate), and consisted of both positive and negative rectangular pulses of random heights and widths. Furthermore, the pulses came in ' showers ' which were separated by apparently random intervals: sometimes by as long as several minutes, and sometimes by less than a second.

Each of the samples shown in figure 1 is a record of a portion of a ' shower '. It will be observed that the rise time is extremely short compared with the pulse widths. The decays are usually sharp, but occasionally exponential. Sample A shows a complex wave which builds up and falls away in steps. Sample B shows a particularly long pulse, while C shows some faster pulses and D pulses of both signs. The electronic equipment was shielded electrically and magnetically, and fixed wire-wound resistors were used in the electrical circuit.

McKay (1954) has suggested that somewhat similar noise in silicon p–n junctions is due to the unstable onset of breakdown in various regions of the junction, where the breakdown is essentially bi-stable and can be switched on and off by thermal fluctuations. An alternative suggestion that the effect may be due to cosmic rays is being investigated.

The authors are grateful to Dr. T. S. Moss for his helpful comments.

References

McKay, K. G., 1954, *Phys. Rev.*, **95**, 887.

A Note on the Structure of Evaporated Silver Deposits Prepared at High Temperatures

By F. H. HERBSTEIN†
Applied Physical Chemistry Laboratory, Imperial College, University of London

Communicated by H. Wilman; *MS. received* 29*th October* 1956

Relatively few investigations have been made of the structures of thin metallic films deposited on polycrystalline substrates at high temperatures, or deposited at room temperature and then annealed at similar high temperatures. The author has recently investigated the structure of silver films prepared in these two ways by electron diffraction. Particular attention has been paid to the effect of using an obliquely incident vapour stream.

---

† Now at the National Physical Research Laboratory, Council for Scientific and Industrial Research, Pretoria, South Africa.

The films were prepared by evaporating silver from a tungsten filament on to a silica plate contained in an evacuated ($\sim 10^{-4}$ mm Hg) silica tube. The assembly could be held at temperatures of up to 900°c. Similar apparatus has been described previously (Brück 1936, Evans and Wilman 1952). The substrate was cleaned by following the procedure recommended by the latter authors. The silica plate was 0·5 cm below the silver source and the deposits were formed with the vapour stream incident at angles from 90° to about 30° to the plate. The rate of deposition was 30–50 Å min^{-1} (at normal incidence) and the time of deposition was held constant at 10 minutes. Thus the final mean deposit thickness (at normal incidence) was of the order of 300–500 Å. All deposits were examined by reflection in a Finch-type electron-diffraction camera, with 50–70 kv electrons and a camera length of about 50 cm.

Deposits were prepared at 20°c, 340°c, 416°c, 595°c and 710°c. There were no signs of orientation in the deposits made at the three lower temperatures. The deposit made at 595°c consisted of large crystals, of which a small proportion was in one-dimensional orientation about the [111] axis. The orientation axis was normal to the substrate even where the incident vapour stream had been inclined to the substrate. There were extra lines on the pattern from this deposit indicating the presence of a small amount of hexagonal silver. This probably resulted from the presence of stacking faults and twinned regions in the normally face-centred-cubic silver crystals. In the deposit made at 710°c (figure 1, Plate) only very little randomly oriented silver remained; the predominant mode of orientation was with the (111) planes parallel to the substrate surface but again without azimuthal orientation. (100) and (110) orientations were present in lesser amounts. There were no signs of the presence of hexagonal silver. In both the deposits showing orientation, the orientation axis was normal to the substrate surface irrespective of the angle of incidence of the vapour stream. Streaks through the (111) diffraction spot indicated the presence of (111) and (311) facets on the crystallites.

These results are in general agreement with previous work in this laboratory and elsewhere (Brück 1936, Rüdiger 1937). Brück found that silver evaporated on to quartz at 240°c showed no orientation; (100) and (111) orientations were found with gold evaporated on to quartz at 480°c, and (110) orientation on quartz at 700°c. Rüdiger found that silver deposited on fused quartz showed orientation only at substrate temperatures above 550°c; (111) orientation was found. Neither of these authors studied the effect of oblique incidence of the vapour stream.

The effect of annealing on deposits made at 20°c was also studied. Heating the Ag deposits *in vacuo* up to 600–700°c produced marked changes in their appearance: the original blue colour of the film, as seen by transmission, changed to milky-white, while the apparent extent of the deposit was reduced. These changes suggest that aggregation (and possibly also partial evaporation) of the film took place on heating. Heating an initially randomly orientated deposit (prepared at 20°c) at 345°c for 40 minutes *in vacuo* produced no orientation, but strong one-dimensional orientation was present after heating such deposits at 610°c and 710°c for 10 minutes. These deposits were oriented mainly with the (100) planes parallel to the substrate, but some (110) and (111) orientation was also found (figure 2, Plate). The orientation axis was normal to the substrate irrespective of the original angle of incidence of the vapour stream. There is also evidence of the presence of a small proportion of hexagonal silver at both temperatures.

These results are, in general, similar to those obtained (Dixit 1933) in experiments on the structure of Ag films deposited on quartz and polished molybdenum at room temperature and then heated. However, in Dixit's experiments only (111) orientation was found for Ag films, although (100) and (110) orientations were found for Al films treated similarly. Swamy (1934) found, by x-ray diffraction, that (111) orientation developed in Au films deposited on quartz and heated to 800°c.

The following conclusions can be drawn from the available results on Ag deposits prepared or annealed at above about 550°c: (*a*) one-dimensional orientation is found, whereas below about 550°c no preferred orientation is observed in thin deposits, at least for annealing times of less than 1 hour; (*b*) the orientation planes are parallel to the substrate surface irrespective of the angle of incidence of the vapour stream, whereas the surface regions of thick deposits of high-melting metals, on substrates initially at room temperature, develop oblique orientation (Evans and Wilman 1952, Murbach and Wilman 1953); (*c*) the nature of the substrate is not important, provided that it is highly smooth; (*d*) the only orientations found are (111), (100) and (110).

In the present investigation (111) orientation has been found to be the predominant mode of orientation in deposits prepared at 600°c and 700°c, while (100) orientation was most marked in room-temperature deposits annealed at these temperatures. Dixit (1933), on the other hand, reports no difference in the structures of annealed deposits from those of deposits freshly prepared at the temperature concerned. A decision on this question, as well as a more quantitative study of the variation of the kind and amount of orientation with deposition (or annealing) temperature, will require a much closer control of the deposition conditions than has been possible with the simple apparatus used until now.

ACKNOWLEDGMENTS

I should like to express my thanks to Dr. H. Wilman for suggesting this problem and for his continued help and encouragement during the course of the work. I am also grateful to the South African Council for Scientific and Industrial Research for enabling me to take part in this research, and for permission to publish this paper.

REFERENCES

BRÜCK, L., 1936, *Ann. Phys., Lpz.*, **26**, 233.
DIXIT, K. R., 1933, *Phil. Mag.*, **16**, 1049.
EVANS, D. M., and WILMAN, H., 1952, *Acta Cryst., Camb.*, **5**, 731.
MURBACH, H. P., and WILMAN, H., 1953, *Proc. Phys. Soc.* B, **66**, 905.
RÜDIGER, O., 1937, *Ann. Phys., Lpz.*, **30**, 505.
SWAMY, S. R., 1934, *Proc. Phys. Soc.*, **46**, 739.

Domain Patterns on Cobalt Crystals

By E. O. HALL

Department of Physics, Sheffield University

MS. received 20th July 1956, *and in revised form 9th November* 1956

THE magnetic domain structure of cobalt crystals, using electropolished specimens, was first described by Elmore (1938), who showed that the specimen consisted of domains magnetized parallel and antiparallel to the single easy direction of magnetization, which is the hexagonal axis. The structure has been confirmed in later work by Mee (1950) and Andrä (1954). However, no attempt has yet been made to examine systematically the patterns formed on crystal planes other than the basal and prism faces, and in consequence the work here described was undertaken.

The specimen used was an ingot of cobalt prepared by the method of Myers and Sucksmith (1951). It contained several large grains, and by sectioning the ingot a sufficient variety of crystal planes was exposed. The ingot was mounted in Woods metal, mechanically polished and then, finally, deeply electropolished in a solution of orthophosphoric acid (S.G. 1·75) in an equal volume of water. The polishing voltage was 1·15–1·2 v. The colloid was prepared in the manner described by Elmore (1938). After microscopic examination, the orientation of selected faces could be determined using standard x-ray back-reflection methods.

It is evident from the work of Sucksmith and Thompson (1954) that the anisotropy constants of cobalt are dependent only on the angle between the direction of magnetization and the *c*-axis. This result is confirmed from figure 1 (*a*) (Plate), which shows the well-known domain pattern on a basal plane. The hexagonal stars show no preferred directions of alignment, confirming the isotropy of magnetic properties in the plane. It would appear that the pattern observed on this plane is only indirectly related to the domain pattern underlying it, since, from electron diffraction studies by Germer (1942), it is known that strong local fields exist at the surface of the basal plane. These fields must arise from the free pole distribution on the domain ends, and the large resulting flux will cause a redistribution of the colloid which will mask the true position of the domain walls.

It is found that this type of pattern is not only a characteristic of the basal plane. Let χ be the angle between the hexad axis and the normal to the plane of observation; then it is found that similar patterns exist on all planes with χ ranging from 0° (the basal plane) to at least 54°. Figure 1 (*b*) (Plate) is an example of the latter case. Between 54° and 62° (figure 1 (*c*), Plate) the pattern changes, with the domain boundaries themselves becoming apparent. The fact that areas of the surface are covered with colloid shows that a free pole distribution is still present on this surface, while the fine scale of the pattern suggests that one is viewing a section of a complex field of main domains containing ‘ dagger ’ domains of reverse magnetization of the type observed by Mee (1950) and Andrä (1954). As χ increases still further the pattern becomes coarser, but a free pole distribution must still be present (figure 1 (*d*), Plate). The pattern finally broadens out to become the familiar prism face pattern.

A theoretical explanation of the patterns may be obtained by considering a simplified domain structure. Assume that a cobalt crystal of unit cross section and length L parallel to the hexad axis is subdivided into domains, each with a square cross section of side D. If the crystal is sheared over through an angle $90° - \chi$, the volume of the crystal and the domains remains constant. Following the treatment given by Kittel (1949), the total energy of the crystal is made of up two parts: (i) the wall energy, and (ii) either the magnetostatic energy, if there is a free pole distribution on the faces, or the ansiotropy energy, if closure domains are formed. In this particular case, the wall energy is $2\sigma_w L/D$, and the magnetostatic energy per unit surface area from a pole distribution of this type is $1{\cdot}06\, I_s^2 D$, where σ_w is the energy of a Bloch wall and I_s the surface pole density. Minimizing the energy then gives the energy per unit volume W_1 in this case as

$$W_1 = 2\left[\frac{2{\cdot}12\, I_s^2 \sigma_w}{L}\right]^{1/2}.$$

In the oblique case, with the crystal sheared over, the relation is simply

$$W_1 = 2\left[\frac{2{\cdot}12\, I_s^2 \sigma_w \cos^2\chi}{L}\right]^{1/2} \text{erg cm}^{-3}. \qquad \ldots\ldots(1)$$

If closure domains were formed, the anisotropy energy involved is $\frac{1}{2}KD$, where K is the anisotropy coefficient, while the wall energy is still approximately $2\sigma_w L/D$. Minimizing the total energy as before, we find the energy per unit volume W_2 is:

$$W_2 = 2\left[\frac{K\sigma_w}{L}\right]^{1/2} \text{erg cm}^{-3}. \qquad \ldots\ldots(2)$$

In the oblique case the volume of the closure domains, as well as that of the crystal, remains constant, while K is a function of χ given by

$$K = k_0 + k_1 \sin^2\chi + k_2 \sin^4\chi.$$

To prove that the free pole distribution is energetically favoured, it must be shown that $W_1 < W_2$ over the whole range of χ. Values of K were thus calculated, using values of $k_0 = 6{\cdot}3$, $k_1 = -7{\cdot}18$, $k_2 = 0{\cdot}95 \times 10^6$ erg cm^{-3} taken from Sucksmith and Thompson (1954), and the ratio W_1/W_2 finally plotted against χ as shown in figure 2. At all angles the ratio is less than unity. The change in the pattern in

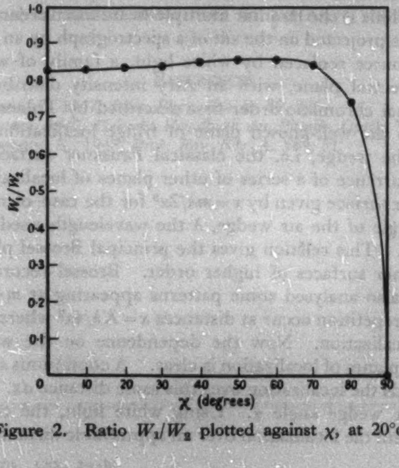


Figure 2. Ratio W_1/W_2 plotted against χ, at 20°c.

the vicinity of $\chi = 60^\circ$ may be due simply to a reduction of the normal flux at the surface, thus allowing the domain structure to be less obscured by colloid agglomerations in the liquid, although it is interesting to note that the theoretical results show that, if the anisotropy were smaller, closure domains would tend to form at about this orientation.

ACKNOWLEDGMENT

The author wishes to thank Professor Sucksmith and Dr. Street for their interest and advice during the course of this work and Dr. Crangle for making available the cobalt ingot.

REFERENCES

ANDRÄ, W., 1954, *Ann. Phys., Lpz.*, **15**, 135.
ELMORE, W. C., 1938, *Phys. Rev.*, **53**, 757.
GERMER, L. H., 1942, *Phys. Rev.*, **62**, 295.
KITTEL, C., 1949, *Rev. Mod. Phys.*, **21**, 541.
MEE, C. D., 1950, *Proc. Phys. Soc.* A, **63**, 922.
MYERS, H. P., and SUCKSMITH, W., 1951, *Proc. Roy. Soc.* A, **207**, 427.
SUCKSMITH, W., and THOMPSON, J. E., 1954, *Proc. Roy. Soc.* A, **225**, 362.

The Formation of White Light Fringes of Equal Chromatic Order

BY N. BARAKAT
Heliopolis University, Cairo, Egypt

Communicated by S. Tolansky; MS. received 22nd December 1955

WITH parallel monochromatic light and a silvered air wedge, Fizeau fringes of equal thickness having a Fabry–Perot intensity distribution can be formed on the Feussner surface of localization close to the interferometer. This is the familiar multiple-beam interference pattern. If the Feussner surface is projected on the slit of a spectrograph by an achromat and the monochromatic source replaced by white light, a family of white-light fringes appear on the spectral plane, with an Airy intensity distribution. They are the fringes of equal chromatic order first described by Tolansky (1945).

In addition to the well-known plane of fringe localization which coincides effectively with the wedge, i.e. the classical Feussner surface, Brossel (1947) predicted the occurrence of a series of other planes of localization at distance x from the Feussner surface given by $x = m\lambda/2\alpha^2$ for the case of normal incidence, where α is the angle of the air wedge, λ the wavelength used and m takes the values 1, 2, 3, ... This relation gives the principal Brossel planes only, which are in fact Feussner surfaces of higher order. Brossel recorded a number of these planes and also analysed some patterns appearing at $m-\frac{1}{2}$. He showed that the planes of repetition occur at distances $x = K\lambda/4\alpha^2$ where K is the order of the surface of localization. Now the dependence on the wavelength of the distance x of such planes of localization is clear. A continuous spectrum between λ and $\lambda + d\lambda$ extends the localization over this same distance dx. Such a distance depends upon the wedge angle α. Using white light, the continuous visible spectrum can extend the localization over an appreciable distance which increases with the order of localization. Thus if any of the Brossel planes F_1, F_2, F_3, ... are projected on a spectrograph slit but using white light, the resulting white light fringes seen on the spectral plane will only be in focus over a limited area, depending upon the extension of localization in space and the depth of focus of the projecting achromat. In fact, strictly speaking, it will not be possible to bring the whole system simultaneously into focus for any other than the zero order of localization. Furthermore as a result, patterns with spacing differing from the original can be expected to be seen in the regions of the spectral plane away from those in focus. They are the white light fringes corresponding to the monochromatic patterns found by Brossel at fractional orders.

Using monochromatic parallel light ($\lambda = 5461$ Å), and a particular set up consisting of a thin silvered air wedge, straight line sharp fringes parallel to the edge of the wedge were easily located on surfaces of localization which for one particular example occurred at distances 9·3, 18·6, 28·2 and 37·5 cm from the normal zero order Feussner surface of localization F_0 which is effectively on the wedge. These are the surfaces of localization F_1, F_2, F_3 and F_4 for this particular wavelength. The spacing of the fringes in such localizations is, as shown by Brossel, similar to F_0, and these indeed do occur in the predicted positions. Considering a visible range, say $\lambda = 4500$ Å to $\lambda = 6500$ Å, the extensions of localization in space for $\Delta\lambda = 2000$ Å are 0, 3·25, 6·5, 9·75 and 13 cm for the orders of localization 0, 1, 2, 3 and 4. It is apparent that the depth of focus of the projecting lens is a minor factor compared with this.

The zero order white-light fringes of equal chromatic order are shown in the Plate at strip *a*. At a distance of 4·6 cm from the previous position sharp Fizeau fringes of half the spacing appear corresponding to $K = \frac{1}{2}$ (for λ 5461 Å). The corresponding white-light fringes are as shown in the Plate at strip *b*. These latter exhibit alternations in intensity so that the true period is the same as the period on F_0. Fringes on F_1 for λ 5461 Å corresponding to $K = 1$ appear at a distance of 9·3 cm from F_0 and the associated white light fringes are as shown on the Plate at strip *c*. Thus only for the zero order Fuessner surface can fringes of equal chromatic order over an extensive wavelength range be secured in focus.

REFERENCES

BROSSEL, J., 1947, *Proc. Phys. Soc.* B, **59**, 226.
FEUSSNER, 1947, *Gehrcke Handbuch der Physik, Optik*, Vol. I.
TOLANSKY, S., 1945, *Proc. Roy. Soc.* A, **184**, 51.

LETTERS TO THE EDITOR

A Single Crystal Photodiode of Lead Sulphide

Lead sulphide cells have in the past made use of the photoconductive properties of microcrystalline layers, prepared either by evaporation (Starkiewicz *et al.* 1947) or by chemical deposition (Kicinski 1948). The cell reported here is photovoltaic, and consists essentially of a p–n junction lying parallel to and a short distance below the surface of a single crystal. An e.m.f. is generated across the junction when the surface is illuminated, the device being thus similar in principle to the silicon solar battery (Chapin *et al.* 1954).

Junctions in lead sulphide have been studied previously by a number of authors (Brebrick and Scanlon 1954, Moss 1956, Bhide *et al.* 1956), but no appreciable photosensitivity has been reported; this was probably due to the width of the junction region being greater than the diffusion length of the generated carriers. We have been able to produce sharp concentration gradients by variation of the stoichiometric balance of lead sulphide, and have obtained junctions of useful sensitivity by converting the surface of a p-type crystal, i.e. one containing excess sulphur, to n-type by vacuum heat treatment. The conversion is caused by the evaporation of sulphur from the surface region, which we found to be several microns thick.

The crystals used in these experiments were grown by the Stockbarger method and were initially n-type; they were converted into p-type by heating in sulphur vapour for a time long enough to ensure a sufficiently uniform penetration of sulphur. The crystals then had an acceptor concentration of about 10^{18} cm^{-3}. Cleaved samples of such crystals, usually of dimensions 8 mm × 5 mm × 1 mm were heated in a vacuum to temperatures between 300 and 500°c, the development of photo-sensitivity being monitored by means of probe electrodes. Figure 1 illustrates how the photosensitivity at room temperature develops after successive heat treatments. In order to make useful photodiodes the treatments were stopped in the region of optimum sensitivity, and permanent electrodes of evaporated gold were then applied. The p-type region was exposed by grinding

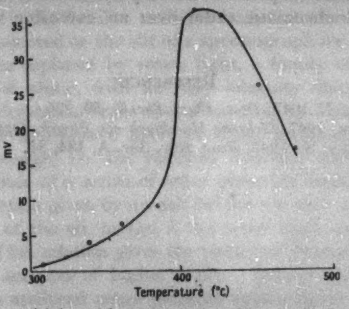


Figure 1. Variation of sensitivity with temperature of successive heat treatments (radiation intensity = 0·05 watt cm^{-2}).

away part of the crystal surface, so ensuring a low resistance contact. The electrode on the n-type surface was in the form of a fine grid; this combined the requirements of high optical transmission and low electrical resistance. Usually the cell was mounted in an evacuated envelope (figure 2), but some preliminary experiments suggested that it may be possible to dispense with this and yet maintain adequate stability.

Figure 2 gives the spectral response of a typical cell of active area 8 mm × 1 mm. The ordinate, the equivalent noise input, is defined as the radiation power in watts required to give a signal equal to the noise of the cell in a band width of 1 c/s.

Measurements were made of the conversion efficiency for total radiation from a Nernst filament source; this was found to be 0·3% for an incident intensity of 0·1 watt cm^{-2}.

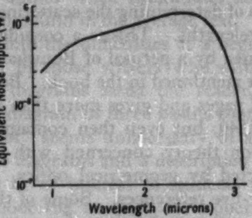


Figure 2. Spectral sensitivity in PbS photodiode. Area = 8 mm^2.

A characteristic feature of the photodiodes is their high speed of response; it is possible to interrupt the radiation at frequencies up to 100 kc/s without a measurable loss of response. The time constant is thus less than 1 microsecond, as compared with values of about 100 microseconds for conventional photoconductive lead sulphide cells.

These cells may be useful in the study of rapid fluctuations of temperature in the range 150–600°c, or in applications where the steady current drain of a photoconductive cell would be undesirable. Under conditions of good illumination sufficient power is available from the photovoltaic cell to operate a relay directly.

Acknowledgment is made to the Admiralty for permission to publish.

Services Electronics Research Laboratory,
Baldock,
Herts.
28th November 1956

J. Starkiewicz.
G. Bate.
H. Bennett.
C. Hilsum.

Bhide, V., Das, J., and Khandekar, P., 1956, *Proc. Phys. Soc.* B, **69**, 245.
Brebrick, R., and Scanlon, W., 1954, *Phys. Rev.*, **96**, 598.
Chapin, D., Fuller, C., and Pearson, G., 1954, *J. Appl. Phys.*, **25**, 676
Kicinski, F, 1948, *Chem. & Ind.* (Rev.), **17**, 54.
Moss, T., 1956, *Proc. Phys. Soc.* B, **68**, 697.
Starkiewicz, J., Sosnowski, L., Simpson, O., 1947, *Nature, Lond.*, **159**, 818.

REVIEWS OF BOOKS

Scattering and Diffraction of Radio Waves, by J. R. MENTZER. Pp. viii + 134. (London and New York: Pergamon Press, 1955.) 30*s*. (Electronics and Waves No. 7.)

The editor of the Pergamon Science Series on Electronics and Waves states that wherever it is practical the monographs will be kept short in length. The one under review is certainly brief, but it is to be doubted whether the subject could be adequately treated in so few pages. In consequence, perhaps, although a book on the scattering of radio waves deserves a welcome, the reviewer finds that his criticism tends to be adverse.

It is said that the aim of the book is to bring research workers and instructors up to date in the methods of determining the scattering and diffraction behaviour of obstacles at radio wavelengths. This is a considerable task; some idea of its magnitude can be gained by a perusal of Bouwkamp's review article (1954), which, incidentally, is not mentioned in the book. Bouwkamp's article occupies over sixty closely packed pages and gives more than five hundred references to papers published since 1940, but even then contains no account of the latest development of diffraction theory concerned with asymptotic corrections to the cross sections predicted by geometrical optics. The book, on the other hand, contains only forty-five references. Some of the omissions (and some of the inclusions, too) seem rather surprising; for example, the behaviour of the field at a sharp edge is mentioned only in connection with the circular aperture and the references cited are Bethe, H. A., 1944, *Phys. Rev.*, **66**, 163, and Houston, R. E., and Noble, R. H., 1951, *J. Appl. Phys.*, **22**, 1295.

In the second chapter various methods of attack on the boundary value problem are discussed, including classical, integral equation and variational methods. All this is in keeping with the stated aim of the book. The discussion on uniqueness is new to the reviewer, though as yet he has to confess inability to master some of the argument.

But when, in the second and third chapters, eleven particular problems are treated, in only one case is a modern method adequately illustrated. Otherwise it is all classical orthogonal function expansions, or assumed current distributions. Whilst these are, of course, important methods, they are scarcely designed to bring the reader up to date. The eleven problems are dealt with in fifty-four pages, and the treatment is sometimes very sketchy; the intelligent reader who does not already know the answer is likely to be puzzled by the absence of any explicit reference to a radiation condition in the account of Macdonald's method for the half-plane.

It is only too easy to be biased in assessing the impact on others of a book on a subject about which one has preconceived ideas of presentation; but it must be said that this attempt, worthy in itself, to expound the interesting work on diffraction of the last decade seems to leave room for improvement.

P. C. CLEMMOW.

Descriptive College Physics, by H. E. White. Pp. x + 485. (New York: Van Nostrand; London: Macmillan, 1955.) 42*s*.

This well-written book is a survey of Physics based on a course given by the author at the University of California for " students not majoring in any of the physical or life sciences ". The text is consequently largely non-mathematical, demanding only a knowledge of simple algebra.

It is an attractive book to read. Beginning with the simple fundamental concepts of mechanics, the text includes clearly written accounts of the properties of matter, heat, sound, light, electricity and magnetism and, finally, atomic and nuclear physics. Perhaps the outstanding merit of the work is that it shows how a very considerable knowledge of physics can be gained with little mathematical equipment.

Much of the material included can be found in the usual physics textbooks for schools, though the author often illuminates aspects of the subject with comments and examples not found in the average text. On the whole, the author displays excellent insight into the difficulties of a beginner and avoids many of the *faux pas* often taken by those writers who write popular expositions of science. Many excellent diagrams assist the reader, but lapses occur occasionally. For example, a diagram of three basins containing water and labelled ' cold ', ' warm ', and ' hot ' seems unnecessary to a reader who is going to persist until he encounters diagrams and a description of the bevatron !

Is this book likely to be of interest to students and teachers in Britain ? Much is said nowadays about the necessity for a liberal education of scientists and technologists. Less is said about the need for educating the student of arts and commerce in science, though this would appear to be just as desirable and might even be profitable ! If courses for this latter purpose existed to any significant extent in Britain, this book would serve admirably to give an outline of physics. It lacks, indeed does not pretend to possess, the more rigorous approach necessary in school and university teaching for the science student. At the same time, it could well be read with profit by physics masters since they would be sure to find many ideas, illustrations and examples in this text which would be helpful in compiling lectures. J. YARWOOD.

Errors of Observation and their Treatment, by J. Topping. Pp. 119. (London: Institute of Physics (Monographs for Students), 1955.) 5*s*.

Dr. Topping's monograph is compact and competent within the limits aimed at and should help academic students who want to acquire a knowledge of the simpler methods of treating observational data. No attempt is made to go beyond the needs of the kind of laboratory work that students of the physical sciences normally encounter. The contents will appear thin to the professional statistician, but I suspect that Dr. Topping takes a very realistic view of what students, already burdened with the learning of basic science, are likely to be able to manage. It would be something achieved if all students of science did have an acquaintance with the simple theory of errors, least squares regression, and the common distributions.

Dr. Topping's book will be a real help to both students and teachers, both for its clear explanations and for the many numerical examples. Some teachers who are not themselves familiar with statistical methods and may suffer from lack of confidence will regret that no answers are given. The references for further reading struck me as a rather odd assortment. They vary considerably in scope and standard of treatment. It is a pity Dr. Topping does not help the student by indicating (a) parallel and (b) next step reading. Some of the references given are moderately advanced in style and the impression gained is that the bibliography as a whole is something of a random sample. M. J. MORONEY.

Mise en équations et résolution des réseaux électriques en régime transitoire par la méthode tensorielle, by A. KAUFMANN. Pp. 104. (Paris: Publications scientifiques et techniques du Ministère de l'Air, N.T. no. 56, 1955.) 1100 *fr*.

With the ever-increasing complexity of electric networks required for use in numerous applications, it became imperative twenty or thirty years ago to systematize the whole process of network analysis, which had hardly progressed since Maxwell's introduction of fictitious ' circulating currents '. As a result, a treatment combining topology and matrix algebra was evolved by both mathematicians and engineers. Prominent among the latter was Gabriel Kron, and the work under review is offered to engineers and physicists by a mathematician as a rigorous and generalized presentation of Kron's so-called ' tensor ' method.

Ths subject is introduced through the surprising medium of abstract vector spaces. These provide a powerful means of obtaining quickly all the basic results, though it is doubtful if more than a small proportion of engineering or physics students in this country (or in France either, judging by the Preface) would make much of the first dozen pages, in spite of their (on the whole) clear and logical style. If, however, the reader can somehow get through them, he will be rewarded by finding little difficulty in following the remaining chapters. These deal with the mesh, node, and mixed mesh–node bases for linear networks of the most general kind, i.e. involving, for example, valves as well as coils, condensers and resistors, with arbitrary initial and applied currents and voltages, Laplace transforms being used to convert differential equations into algebraic ones. Several weighty illustrative examples are carefully worked out.

A valuable bibliography with more than a hundred entries completes the book. A. TALBOT.

Les fonctions orthogonales dans les problèmes aux limites de la physique mathématique, by T. VOGEL. Pp. iv + 191. (Paris: Editions du Centre National de la Recherche Scientifique, 1953.) 1200 *fr*.

This monograph is written for physicists and engineers who have occasion to deal with the solution of what are loosely known as ' the equations of mathematical physics '.

The first chapter gives an account of the general theory of orthogonal functions. It begins with an account of the basic properties of functional spaces and of the concept of orthogonality in such spaces, goes on to discussions of the theory of systems of linear differential equations (ordinary and partial), of the closure of certain sequences of functions and concludes with a brief discussion of classical perturbation theory.

The second chapter is devoted to the study of the properties of those sets of orthogonal functions which occur most frequently in applications to physics—trigonometric functions and Fourier series, Bessel functions, Legendre functions, Mathieu functions and the polynomials of Jacobi, Tchebycheff, Laguerre, Hermite and Charlier.

The application of these functions (and their associated series expansions) to the solution of boundary value problems in classical and quantum physics is considered in the final chapter. The problems considered are taken from wave mechanics, potential theory, diffusion theory, elasticity and the theory of wave propagation.

There is a sufficiently full bibliography at the end of each chapter to enable the reader to continue his study of the topics raised, and (at the end of the monograph) a very useful table giving the sums of certain Fourier series with simple coefficients. This monograph will make difficult reading for a physicist who has not also studied mathematics extensively. It is beautifully written but it assumes on the part of the reader a high degree of maturity and an acquaintance with such abstract topics as Lebesque integration. (The brevity of the exposition can be judged by the fact that the theory of Bessel functions is dealt with in less than ten pages.) It should prove a valuable textbook for students of applied mathematics or theoretical physics.

I. N. SNEDDON.

The Theory of Games and Linear Programming, by S. VAJDA. Pp. 106. (London: Methuen's Monographs on Physical Subjects, 1956.) 8*s*. 6*d*.

The present vogue for the theory of games dates only from 1944 when a monumental original work on the subject was published by von Neumann and Morgenstern. The subject is by no means as frivolous as the title may suggest since many situations arising in economics and in warfare are illuminated by it. Linear programming is a related but quite distinct subject. Here the object is to maximize or minimize a linear function of a given number of variables subject to constraints expressed as linear inequalities. An example of a practical problem which can be solved by linear programming is that facing a housewife who is able to buy a variety of foodstuffs each containing a different proportion of protein, carbohydrate, etc. and who wishes to provide her family with a balanced diet at minimum cost.

Both the theory of games and linear programming are subjects which may prove to have applications in unexpected places and Dr. Vajda has presented their essentials in an admirable manner. His book can be highly recommended, not only to the general scientific reader who has heard of the subjects and who wishes to know what they are about, but also to the student who is embarking on a serious study.

M. V. WILKES.

CONTENTS OF SECTION A

PAGE

Dr. C. A. ten Seldam. Calculation of the Interatomic Distance at which a Model of Compressed Solid Helium becomes a Metal 97

Mr. T. J. M. Boyd, Dr. B. L. Moiseiwitsch and Dr. A. L. Stewart. Inelastic Collisions between Heavy Particles. Part V: Electron Loss for Fast He$^+$ Ions passing through Atomic Hydrogen, and Ionization of Hydrogen Atoms by Fast He$^+$ Ions 110

Mr. J. Adler and Dr. B. L. Moiseiwitsch. Inelastic Collisions between Heavy Particles. Part VI: The 2s^{3}S–2p and 3p^{3}P Excitations of Fast He Atoms by H and Ne Atoms 117

Dr. T. B. Grimley. Homopolar Effects in Crystals having the Sodium Chloride Type of Lattice 123

Mr. F. W. Bullock. Nuclear Disintegrations produced by Cosmic Ray Particles in a High Pressure Cloud Chamber 134

Mr. V. Knapp. Nuclear Resonant Scattering of Gamma Rays in ^{199}Hg . . 142

Mr. A. J. Apostolakis and Mr. I. Macpherson. An Emulsion Spectrograph—A New Technique for the Determination of the Momenta of Fast Charged Particles. Part I: Design Consideration and Theory 146

Mr. A. J. Apostolakis and Mr. I. Macpherson. An Emulsion Spectrograph—A New Technique for the Determination of the Momenta of Fast Charged Particles. Part II: Construction and Applications of a Normal Incidence Stack 154

Letters to the Editor:

Mr. C. J. Batty and Dr. S. J. Goldsack. Polarization of Protons Scattered from Carbon at 970 MeV 165

Reviews of Books 167

Contents of Section B 168

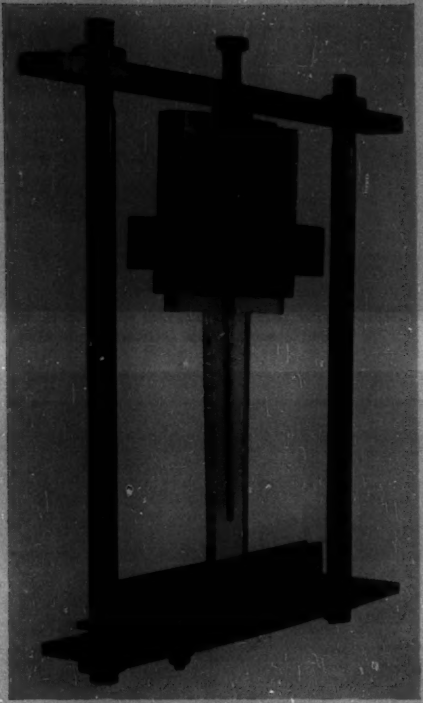


Figure 1. Apparatus for producing controlled fractures.

Time constant of C.R.O. amplifier = 0·5 sec.

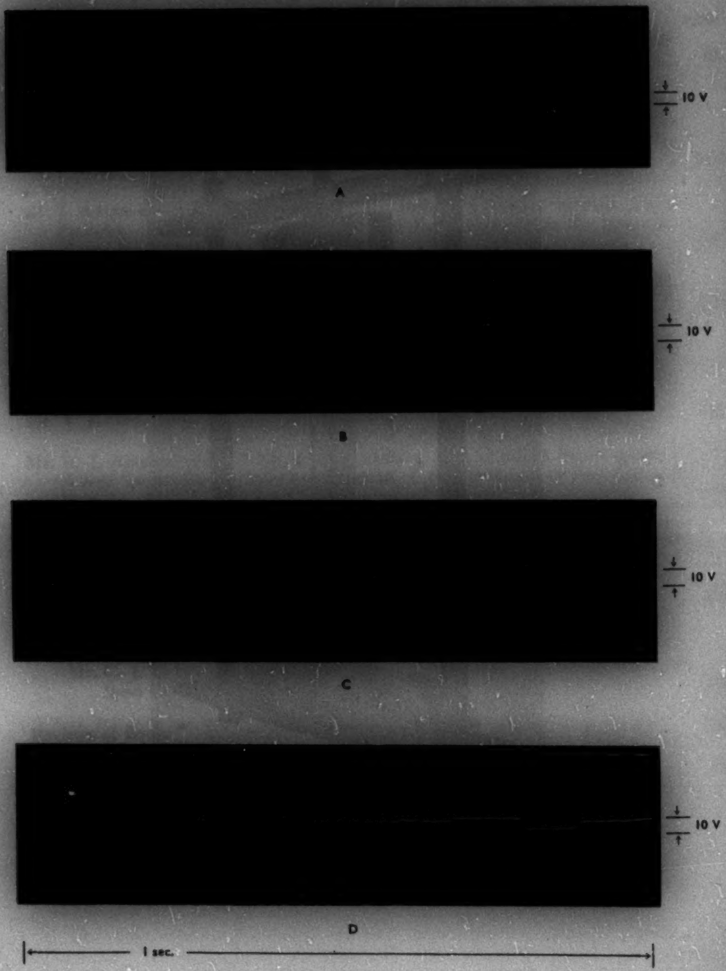
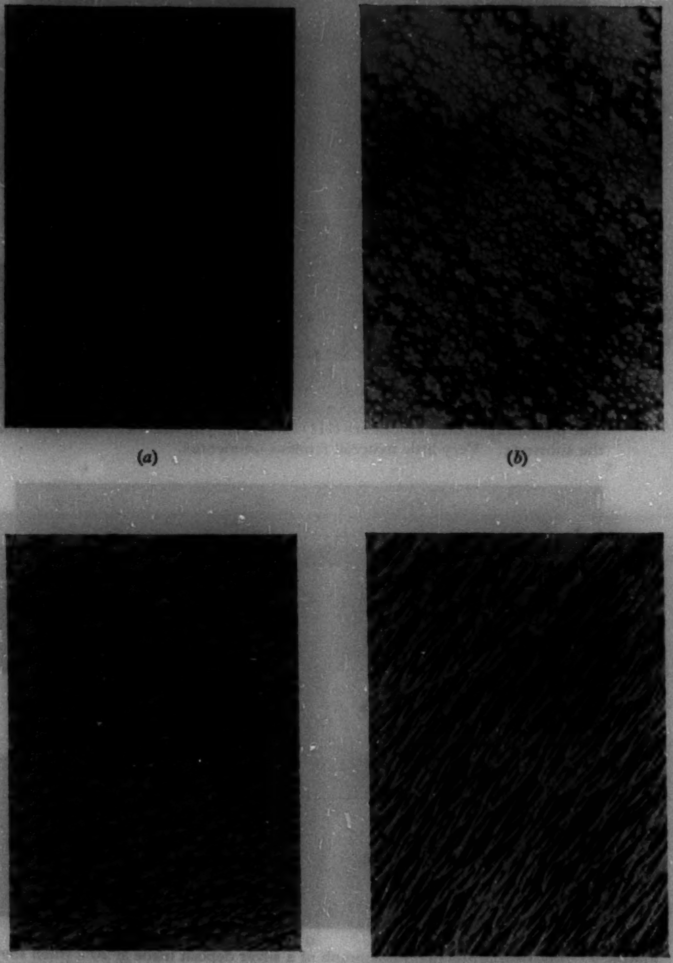


Figure 1. Noise voltage at collector during secondary ionization.

Figure 1. Electron-diffraction photograph of a silver film deposited on silica at 710°c. The vapour stream was incident at an angle of about 35° to the substrate surface. The predominant orientations are (111) and (110), with the orientation axes normal to the substrate. Very little material remains unoriented.

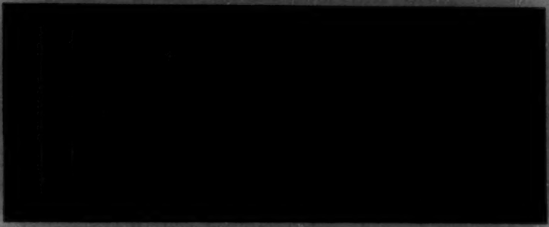
Figure 2. Electron-diffraction photograph of a silver deposit made at 20°c (then showing random structure) and heated at 610°c for 10 minutes *in vacuo*. The vapour stream was incident at an angle of about 35° to the substrate surface. The predominant orientations are (100) and (110), with a trace of (111). The orientation axes are normal to the substrate. An appreciable proportion of the silver deposit remains unoriented.



(*a*) (*b*)

(*c*) (*d*)

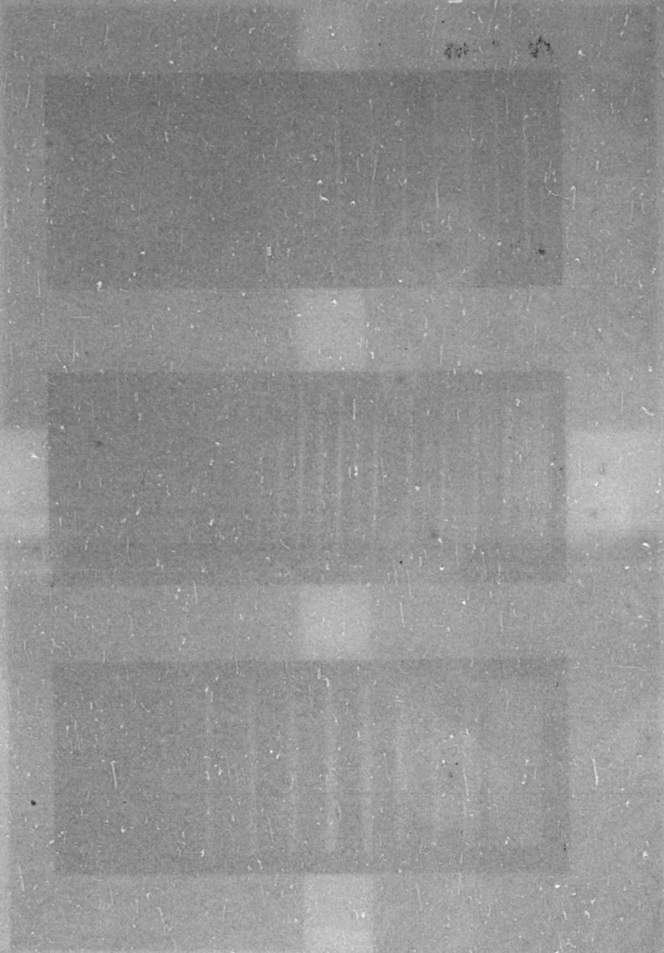
Figure 1. Domain patterns on cobalt crystals (× 150). (*a*) $\chi=0^\circ$ (basal plane); (*b*) $\chi=54^\circ$; (*c*) $\chi=62^\circ$; (*d*) $\chi=76^\circ$.



a

b

c



[illegible]

The Epitaxial Growth and Oxidation of Nickel, Cobalt and Iron on Rocksalt

By L. E. COLLINS[†] and O. S. HEAVENS[‡]
Physics Department, University of Reading

MS. received 13*th August* 1956, *and in final form* 25*th October* 1956

Abstract. The structure and orientations of films of iron, nickel and cobalt, deposited by thermal evaporation on rocksalt, have been examined by electron diffraction. The oxides, forming spontaneously on the epitaxially grown metal films, have also been examined. The results are considered in the light of current theories of epitaxy and it is shown that the lattice misfit between substrate and deposit materials is generally less important than the maintenance of correct co-ordination of atoms in the first layer. Generally no evidence of a strained deposit lattice was obtained except in the case of iron oxide films on iron.

§ 1. INTRODUCTION

A STUDY has been made by electron diffraction of the growth of films of nickel, cobalt and iron on heated rocksalt crystals. The films were deposited from the vapour on the cube, dodecahedral and octahedral faces of crystals at temperatures up to 500°c. The structures of the films were determined from electron diffraction patterns obtained both before and after removal of the film from the substrate. The growth of the oxides on the surface of epitaxially grown films of the metals has also been studied. The evidence for the existence of a hexagonal modification of the structure of nickel in sputtered or evaporated films is considered; it is shown that the reported hexagonal structure could be that of nickel nitride. In the case of cobalt, both cubic and hexagonal forms of the metal are observed; a lamella-type structure is found with stacking faults on {111} planes.

§ 2. EXPERIMENTAL

In order to avoid atmospheric contamination, films were prepared in the specimen chamber of the electron diffraction camera. The axis of this instrument is horizontal and the specimen chamber is large so that the evaporation sources, heating elements and shutters needed for the preparation of films may be easily accommodated. For the convenient examination of reflection patterns it is necessary to arrange for delicate tilting of the specimen about an axis in its own plane. This is effected by the specimen adjuster shown in figure 1. The centre of curvature of the spherical surfaces lies in the plane of the specimen. The 'O' ring between the surfaces remains vacuum tight even while tilting is effected.

In the design of the specimen holder, on which the rocksalt substrate is to be heated to about 500°c, care must be taken to eliminate deflection of the electron

---

[†] Now at Atomic Weapons Research Establishment, Aldermaston, Berks.
[‡] Now at Royal Holloway College, University of London.

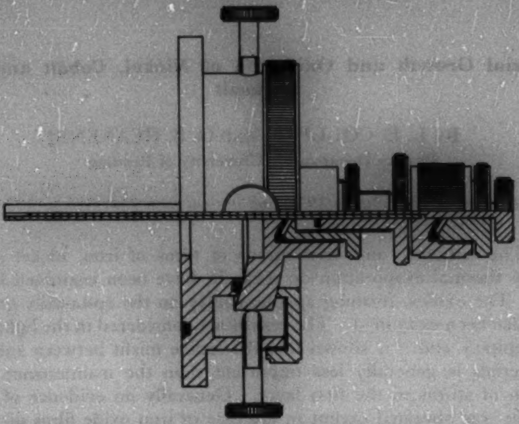


Figure 1. Specimen adjuster. Transverse motion effected by ratchet and pinion operated by inner shaft.

beam by the magnetic field of the heating elements. This may be effected either by a flat heater or by a toroidal heater. The form used in these experiments is shown in figure 2. The rocksalt slab is held against a copper core; the heaters are carried in holders of pyrophillite which have been baked for several hours at 1200°c after forming. The temperature is measured by a platinum : platinum–rhodium thermocouple placed immediately behind the

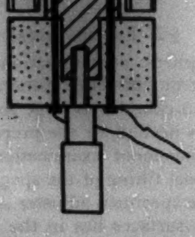


Figure 2. Specimen heater.

rocksalt slab. The surface temperature of the slab will be slightly below the value indicated by the thermocouple.

The films were prepared by evaporation of metal wire from a helix of tungsten. A shutter was interposed between source and target during the first stages of the evaporation; this prevented any surface contamination of the source from reaching the target. Film thicknesses were estimated from the transmittance of the films to visible light and are approximate only.

'Specpure' nickel, cobalt and iron were used in the experiments. None of the electron diffraction plates showed rings or spots characteristic of the tungsten lattice. In the light of the results of tests for tungsten contamination on silver and germanium films (Heavens 1952), it is unlikely that the tungsten content of the films is sufficiently high to influence the structures of the metal films.

§ 3. RESULTS ON THE GROWTH OF NICKEL ON ROCKSALT

3.1. *Dependence of Structure Type on Substrate Temperature*

For substrate temperatures below 350°C polycrystalline films are formed, the crystallite size increasing with increasing temperature. At about 350°C the crystallites show a tendency to form with (111) planes roughly parallel to the substrate plane, this tendency being enhanced by high (~100 Å sec^{-1}) rates of deposition. On (001) rocksalt faces at 370–400°C, (110) and (111) fibre orientation are observed together with some (001) single crystal regions. Deposition at 100 Å sec^{-1} on rocksalt at 450°C produces completely orientated Ni films with (001) planes parallel to the rocksalt cube face. At higher substrate temperatures and rates of deposition quadrupling of high order diffraction spots indicates the development of {111} crystal faces. This behaviour has been observed in cobalt films deposited on single crystals of copper (Cochrane 1936).

3.2. *Monocrystalline Films*

The orientations which were observed in films deposited at 100 Å sec^{-1} and at a substrate temperature of 450°C are given in table 1.

Table 1

Rocksalt	Nickel	Remarks
(001) [100]	(001) [100]	Predominates
(001) [100]	(001) [110]	Very rare, occurring only in very thin layers.
(110) [001]	(110) [001]	Lamellar structure due to repeated twinning
(111) $[\bar{1}10]$	(111) $[\bar{1}10]$	

Films of various thicknesses were examined; in all cases the interatomic spacings were found to correspond, within ±0·5%, to those of the normal Ni lattice. Some aggregation is evident in thin (~20 Å average thickness) films since the spot size indicates a crystallite size of the order of 50 Å. No twinning is observed in films of this thickness. Films of a few hundred ångströms thickness invariably show signs of twinning.

The transmission pattern of a (001) film removed by dissolving away the substrate, is shown in figure 3 (*a*) (Plate I). Considerable distortion about the [100] and [110] directions is in evidence (Goche and Wilman 1939). NiO is present in two orientations, viz. (001) NiO ‖ (001) Ni and (111) NiO ‖ (001) Ni. The former orientation predominates if prolonged heating at 450°c occurs: as the oxide thickness grows, however, the monocrystalline layer degenerates into a polycrystalline one. The twelve low-angle spots in figure 3 (*a*) arise from secondary scattering of the NiO and Ni primary diffracted beams by the Ni and NiO lattices respectively. Figure 3 (*b*) (Plate I) shows the calculated positions of the strongest spots arising from secondary diffraction of beams from Ni in (100) orientation by NiO in (111) orientation.

After heating (001) Ni films freed from the NaCl substrate to 650°c for two hours at a pressure of 10^{-6} mm Hg, all signs of distortion and twinning disappear. Films so treated are found by electron microscopy to be continuous rather than aggregated. Results of experiments on the magnetic properties of such single-crystal films have been reported elsewhere (Collins and Heavens 1954).

Table 2 summarizes the orientations observed of NiO on Ni.

Table 2

Nickel	Nickel Oxide	Remarks
(001) [100]	(001) [100]	Oxidation at 450°c.
(001) [110]	(111) $[\bar{1}10]$	Oxidation at room temperature
(110) [001]	(110) [001]	
(110)	(001)	[110] NiO at ±10½° to $[\bar{1}10]$ Ni†
(111) $[\bar{1}10]$	(111) $[\bar{1}10]$	

3.3. *Hexagonal Structure*

The patterns from several specimens contained rings or spots consistent with a hexagonal structure with $a = 3{\cdot}12 \pm 0{\cdot}03$ Å and $c = 5{\cdot}00 \pm 0{\cdot}05$ Å. This does not correspond with the lattice of hexagonal nickel which has been observed several times in sputtered films (Hull 1917, Bredig and Allolio 1927, Thomson 1931, Bredig and Schwarz von Bergkampf 1931, LeClerc and Michel 1939) for which the lattice parameters are $a = 2{\cdot}6$ Å, $c = 4{\cdot}3$ Å. In this connection we may make two comments: (i) that the hexagonal structure was not observed when sputtering was effected using spectroscopically pure argon and helium (Ingersoll and Hanawalt 1929) (ii) nickel nitride Ni_3N has a hexagonal structure with $a = 2{\cdot}664$ Å, $c = 4{\cdot}298$ Å (Juza and Sachsze 1943). It would appear possible that the hexagonal forms observed in the early sputtering experiments were, in fact, the nitride and not metallic nickel. No hexagonal structure with these parameters has been observed in these experiments on evaporated films.

§ 4. Results on the Growth of Cobalt on Rocksalt

4.1. *Dependence of Structure Type on Substrate Temperature*

In contrast to the case of nickel, there is ample, unequivocal evidence for the existence of both hexagonal ($\alpha-$) cobalt, which is stable (in bulk) at temperatures up to 420°c (Brück 1936), and face-centred cubic ($\beta-$) cobalt which is the more stable form above this temperature.

---

† Although the diffractions from NiO in this orientation lie close to the {111} Ni twin spots, they are nevertheless resolved.

In the present experiments, cobalt films deposited at temperatures below 400°c are found to be mainly polycrystalline. Above 480°c well-defined single-crystal films form whilst between 400°c and 480°c development from fibre orientation to the monocrystalline form takes place. Of considerable interest is the presence of cubic cobalt in films deposited on rocksalt maintained at room temperature, suggesting that the large kinetic energy of the impinging atoms allows the formation of cubic cobalt which is then frozen in by the rapid cooling. As the substrate temperature is increased from room temperature, films are formed containing a steadily increasing proportion of cubic cobalt. Hexagonal cobalt is sometimes observed even in films deposited at 500°c.

4.2. *Structure of Cobalt Films deposited on the* (001) *Face of Rocksalt*

In the study of epitaxy it is clearly an advantage to examine the thinnest films possible in order that features in the layers of atoms close to the substrate are not masked by the diffractions of the overlying metal. In the case of cobalt, it was possible to obtain patterns corresponding to films of average thickness only a few ångströms. Owing to oxidation by the residual oxygen in the camera, such patterns last for a few minutes only: since a diffraction pattern is obtained in an exposure time of 1–2 seconds, the pattern corresponding to the metal itself is easily obtained.

Figure 4 (Plate I) shows the pattern obtained from such a layer deposited on (001) rocksalt at 460°c, immediately after deposition (beam along [110]). Spots from both rocksalt and cobalt are present, those from the former carrying refraction streaks which show that the rocksalt surface is very smooth. The cobalt spots bear no such streaks, showing that the cobalt is in the form of discrete crystallites on the rocksalt surface. On the original negative, Kikuchi lines from the rocksalt are faintly visible which indicates (Heavens 1953) that the average thickness of the cobalt on the surface cannot exceed a few ångströms. Examination by visible microscopy does not reveal any aggregation so that this must, if present, be on a small scale. It is likely, therefore, that the crystallites on the surface are of the order of tens of atoms across.

From figure 4 it is seen that the Co deposits with the (001) plane parallel to the rocksalt in two orientations, viz. [100] Co $\|$ [100] NaCl and [100] Co $\|$ [110] NaCl. It is also seen that the Co lattice is not strained to the rocksalt.

On removal from the substrate, thin (<200 Å) cubic cobalt films are observed to change to a polycrystalline hexagonal form. The cubic structure, which is crystallographically unstable at room temperature, is clearly maintained by the rocksalt influence. It is likely that the stacking faults in the (111) planes act as centres of recrystallization for the formation of hexagonal cobalt. In much thicker films, however, the cubic form is observed even after removal of the film from the substrate.

The main features of interest in the patterns obtained from a Co film 900 Å thick deposited at 20 Å sec^{-1} on rocksalt at 500°c are: (*a*) a strong centred rectangular array from cubic cobalt with (001) Co $\|$ (001) NaCl and [100] Co $\|$ [100] NaCl, (*b*) twin spots on the reciprocal lattice cell diagonals indicating twinning on all octahedral $\{111\}$ faces, (*c*) streaks forming a diamond-shaped network, due to cross grating diffractions from (111) and $(11\bar{1})$ twin planes, showing that successive lamellae are extremely thin and so are acting as two-dimensional gratings.

The transmission patterns (figure 5, Plate II) obtained after removal from the substrate are similar to those obtained for silver by Goche and Wilman (1939) and indicate severe distortions about the [100] and [110] directions. Except in films prepared at temperatures above 550°c, spots from hexagonal Co are present with the orientation (0001) hexagonal ‖ (111) cubic. Streaks through the [111] directions of cubic Co show the presence of stacking faults. Most, although not all, specimens show twin spots.

In specimens in which appreciable oxidation has taken place at high temperatures, the oxide is found to be in parallel orientation on the (001) plane of Co, with [100] oxide parallel to [100] Co. When oxidation occurs at temperatures well below the orientation temperature, the tendency is for the oxide to form with (111) planes parallel to the (001) Co and with the cube diagonals of the oxide and the cobalt parallel. There is evidence of considerable distortion of the oxide film even when little distortion is observed in the underlying cobalt film. This is due to peeling of the oxide from the cobalt, similar to the behaviour observed (Wilman 1940) in the growth of silver chloride films on single crystals of silver.

The orientations of Co and CoO observed in these experiments are summarized in tables 3 and 4. The substrate temperature was maintained at 500°c; the rate of deposition was of the order of 20 Å $\sec^{-1}$.

Table 3

Rocksalt Face	Cobalt Face	Directions Parallel Co	NaCl	Remarks
(001)	(001)	[100]	[100]	Predominates. Twinning. Stacking faults.
	(001)	[100]	[110]	Weak. Occurs only in thin (< 50 Å) layers.
(110)	(110)	[001]	[001]	Stacking faults in (111) planes.
(111)	(111)	$[\bar{1}10]$	$[\bar{1}10]$	(111) twinning.

Table 4

Cobalt Face	Oxide Face	Directions Parallel CoO	Co	Remarks
(001)	(001)	[100]	[100]	Mainly randomly orientated.
(001)	(111)	$[\bar{1}10]$	$[\bar{1}10]$	Oxidation at room temperature.
(110)	(110)	[001]	[001]	
(111)	(111)	$[\bar{1}10]$	$[\bar{1}10]$	

The effect of the presence of stacking faults in materials which may form cubic or hexagonal lattices has been studied by Wilson (1942).

The reciprocal lattice points are elongated in the *c* direction of the hexagonal system, i.e. in the [111] direction of the cubic system. The length of the streaks through the (111) diffractions gives an indication of the density of faults in the crystal (Taylor, Hinde and Lipson 1951). In patterns from cobalt films grown epitaxially at 480°c, short streaks are obtained indicating that the regions of hexagonal and cubic crystals are each several layers thick. Broadening of the hexagonal spots shows, moreover, that cubic material predominates. This was the state of affairs generally found in films deposited at temperatures above 480°c and is to be expected in the light of the known higher stability of the cubic form at this temperature.

§ 5. RESULTS ON THE GROWTH OF IRON ON ROCKSALT

Unlike nickel and cobalt, the structure of iron (at temperatures below 900°c) is body-centred cubic. We should, therefore, expect to observe different orientations when iron is deposited on rocksalt from those obtained for Ni and Co. Results for the (001) rocksalt face have been reported by Brück (1936) and Shirai (1937, 1938), the following orientations being observed: (*a*) $\mathrm{Fe}(001)[100] \parallel \mathrm{NaCl}(001)[100]$ and (*b*) $\mathrm{Fe}(001) \parallel [110] \parallel \mathrm{NaCl}(001)[100]$. Although these were found to be the most popular forms, two other orientations were also found, viz. $(111)\,\mathrm{Fe} \parallel (001)\,\mathrm{NaCl}$ with $[11\bar{2}]\,\mathrm{Fe} \parallel [\bar{1}10]\,\mathrm{NaCl}$ and $(210)\,\mathrm{Fe} \parallel (001)\,\mathrm{NaCl}$ with the four equivalent orientations in which [112] type directions lie along [100] NaCl.

It was found in the experiments to be described that either orientation (*a*) or (*b*) could be obtained and that sometimes a mixture of the two formed. An attempt was made to determine the conditions which favoured one or the other orientation. In the temperature range 400°c to 500°c, no correlation was found between the occurrence of one or the other orientations and the following factors: rate of deposition, substrate temperature, film thickness.

In view of the complexity of the results on the oxidation of the Fe films, we shall deal in this section only with the results of the growth of iron or rocksalt and discuss the oxidation results in § 6.

5.1. *Structure of Fe Films deposited on the* (001) *and* (110) *Faces of Rocksalt*

The results obtained for the (001) face, at a temperature of 470°c and a rate of deposition of the order of $50\,\text{Å}\,\mathrm{sec}^{-1}$ are similar to those quoted by Brück and Shirai. In transmission patterns of films prepared at temperatures below 440°c some rotational disorientation (probably rotational slip) about [001] Fe is present (Evans and Wilman 1950, Wilman 1951). In films grown on (110) rocksalt, two orientations are observed:

$$\mathrm{Fe}(110)[001] \parallel \mathrm{NaCl}(110)[001] \quad \text{and} \quad \mathrm{Fe}(111)[\bar{1}10] \parallel \mathrm{NaCl}(110)[001].$$

5.2. *Structure of Fe Films deposited on the* (111) *Face of Rocksalt*

The transmission pattern obtained is shown in figure 6 (Plate II). There are two readily distinguishable arrays, one of sharp spots and the other of broader arcs. The array of sharp spots corresponds to the parallel orientation of (111) Fe on (111) NaCl. Since both planes contain arrays of three-fold symmetry, the conditions may be expected to allow the extensive growth of this orientation indicated by the sharpness of the diffraction spots. The broader arcs arise from the orientation $(110)\,\mathrm{Fe} \parallel (111)\,\mathrm{NaCl}$, with the cube edge directions of Fe along the face diagonals of NaCl. It is seen that these three orientations are those in which principal rows of atoms in the (110) Fe planes coincide in direction with the principal rows in (111) NaCl planes. A third, weaker, orientation is present, namely $\mathrm{Fe}(111)[\bar{2}11] \parallel \mathrm{NaCl}(111)[\bar{1}10]$.

§ 6. RESULTS OF THE FORMATION OF IRON OXIDE ON EPITAXIALLY GROWN IRON FILMS

6.1. (001) *Iron Films: Electron Beam along* [100] *Fe*

Oxidation of the iron films occurs rapidly and spontaneously at a temperature of 450°c and pressure 10^{-4} mm Hg. The pattern obtained within one or two seconds of deposition shows spots from the iron lattice only. The (001) Fe plane is parallel to (001) NaCl with [100] Fe ∥ [110] NaCl. Figure 7 (Plate II) was obtained from this film one minute after deposition. The oxide diffractions form a centred rectangular array and correspond to a layer of FeO orientated with its cube face parallel to the cube face of the iron and with the cube face diagonal of the oxide parallel to the cube edge of the Fe lattice. Measurement of the FeO lattice spacings indicates that the FeO lattice is strained to the spacing of Fe in the plane of the deposit to an experimental accuracy within 1%. The lattice spacing of the (110) planes of iron is not known with certainty at the temperature of 400°c but will lie close to the value 4·05 Å. The normal lattice spacing of FeO, from x-ray evidence, is 4·294 Å, so there would appear to be a contraction of 5–6% from the normal spacing in the plane of the film. The suspicion of a streak between the (113) FeO and (002) Fe diffractions indicates the beginnings of the transition, which is discussed below, from FeO to the Fe_3O_4 in spinel structure.

As oxidation proceeds, a broadening is observed in some of the oxide diffractions, sometimes with the formation of a doublet. Since the oxide layer is initially strained to the Fe spacing, and since, on further growth, the normal oxide spacing will be attained, the broadening is simply a manifestation of the varying oxide lattice parameter.

Figure 8 (Plate II) represents the next stage in the oxidation process. Diffractions from the iron lattice are still prominent and many extra spots and rings are seen. There is a marked broadening of the rings at the intersection of [001] Fe through 010 Fe. The oxide ring through the 006 oxide diffraction passes through the 033 Fe spots, indicating that the spacings are the same to within 1%.

The structure of the oxide at this stage is intermediate between those of FeO and Fe_3O_4. As oxidation proceeds, the Fe_3O_4 (spinel) diffractions increase in intensity whilst the spots from the iron and from the intermediate oxide phase diminish. Even at the stage where the transition to Fe_3O_4 is complete (figure 9, Plate II), the iron spots are generally detectable on the negative. The lattice constant of the final Fe_3O_4 structure is found to correspond to the bulk value, suggesting that, by the time the spinel structure has completely developed, the influence of the underlying iron lattice is lost. The two orientations of Fe_3O_4 present in figure 9 correspond to the two Fe orientations initially present.

We may summarize the oxidation of the (001) plane of iron at 450°c and at 10^{-4} mm Hg as follows. Initially FeO is formed with (001) FeO ∥ (001) Fe and with the FeO cube face diagonal parallel to the cube edge of iron. The spacing of the oxide in the plane of the surface is measurably less than that of the bulk material. Within a few seconds a range of structures intermediate between that of FeO and Fe_3O_4 spinel is observed, probably arising from the formation of lattice vacancies in the FeO. There is a continual change of structure at this stage as the lattice vacancies become ordered. In the final stage, Fe_3O_4 is formed in parallel orientation on the original FeO. The lattice spacing of the Fe_3O_4 is found to be normal.

The structures of iron oxide films formed on iron grown epitaxially on LiF have now been studied by Haase (1956) with conclusions generally similar to those given above. An exception is that the Fe_3O_4 films grown on LiF were found to be twinned: no such twinning was observed in the experiments described here.

6.2. *Oxide Films on Iron grown on* (110) *and* (111) *Rocksalt Planes*

The iron orientations in these cases have either the (110) or (111) planes of iron parallel to the plane of the film. On the (110) planes of iron, oxide orientation with (110) planes parallel to the substrate with $[\bar{1}10]$ FeO || [001] Fe is shown by a fairly sharp spot pattern. A system of much broader arcs shows a tendency to the formation of the orientation (111) FeO || (110) Fe with $[\bar{1}10]$ FeO || [001] Fe. The latter orientation tends to occur by low temperature oxidation. Similar behaviour is observed on the (111) Fe planes. Either the (110) or the (111) plane of oxide lies parallel to the octahedral iron planes, in either case with $[\bar{1}10]$ oxide parallel to $[\bar{1}10]$ metal. The lattice spacing of the final Fe_3O_4 structure was found to be normal, within an experimental error of ± 1%.

§ 7. THEORIES OF EPITAXY

Before discussing the results given in the preceding sections, we shall mention briefly the main points of the various theoretical treatments of epitaxy which have appeared. Frank and van der Merwe (1949) have examined the conditions for stability of a two-dimensional array of atoms subjected to a periodic force due to the substrate atoms. Such a system is shown to be unstable for misfits greater than 9% between substrate spacing and monolayer spacing. At low temperatures, a metastable configuration with up to 14% misfit is possible. Smollett and Blackman (1951) studied the stability of two-dimensional arrays of atoms with the configuration of the lattice planes of various alkali halides. Several pairs of alkali halides are considered between which epitaxy is known to occur and it is shown that, for many cases, the strain required to produce a monolayer of the deposit material to fit with the substrate would lead to instability. Conversely it is shown that, in certain cases where a strained layer would be stable, epitaxy does not occur. These treatments lead to the directly conflicting views for and against the existence of a strained contact monolayer. Since precise measurement of the interatomic spacing in the contact monolayer has not so far been achieved, direct, decisive evidence on this point is not available.

Drabble (1949) has examined the problem of stability of deposits in terms of the co-ordination conditions of the deposited atoms and has shown that the observed orientation is frequently that in which there is a continuity of structure across the deposit–substrate boundary. Thus if silver atoms are placed on a (001) face of rocksalt in positions relative to the chloride ions similar to those in silver chloride, then parallel orientation results: {Ag (001)[100] || NaCl (001)[100]}. This is the orientation observed experimentally, despite the large rocksalt–silver lattice misfit of 27·8%.

Engel (1952) has considered the growth of metal deposits on rocksalt on the assumption that the deposit metal is ionized and that the binding to the surface is predominantly electrostatic. In the case of silver and rocksalt, the correct orientation is predicted. If the appropriate ionization state of the metal is assumed, then the epitaxial temperature is found to increase linearly with ionization potential. Examination of results in the light of this treatment is difficult on two grounds: (i) it is impossible to decide which ionization state to assume for the added metal, (ii) values of the higher ionization potentials are in many cases not available.

The evidence for and against the presence of a strained monolayer is conflicting. ZnO on Zn has been studied by Finch and Quarrell (1933, 1934), who reported pseudomorphic growth, and by Acharya (1948), Raether (1950) and Lucas (1951), who found no evidence of pseudomorphism. Uyeda (1942) found no evidence of pseudomorphism in silver layers on molybdenite. Menzer (1938 a, b) suggested that large misfits may be accommodated by the development of twinning in layers close to the surface. However, in certain of the experiments discussed below, no twinning was observed even when there existed large lattice misfits. Evans and Wilman (1950) suggest that the rotational slip observed in ZnO films on Zn serves as a means of releasing the strain attendant on initial (pseudomorphic) growth. Similarly, the presence of twinning, which can be removed by annealing, has been assumed (Goche and Wilman 1939) to be a means of releasing strain in the initially formed layers. It is clear, however, that strain is to be expected in a film as a whole, whether or not pseudomorphic growth occurs and that the mere presence of strain is not unequivocal evidence of pseudomorphism.

§ 8. THE GROWTH OF OXIDE FILMS

The laws governing the growth of oxide films on metal surfaces depend, at least in the early stages of growth, on the structure and orientation of the oxide formed. Different conditions obtain depending whether or not epitaxial growth occurs. It is likely, therefore, that a study of the epitaxial growth of oxides on metals will be able to contribute to theories of oxidation generally. Although the mechanism of formation of oxide films differs from that governing the growth of films deposited from the vapour, similar conditions relating to stability, misfit, etc., may be expected to apply. Changes may be expected in the rate of oxidation of a surface if, for example, the mode of growth changes from epitaxial to one in which a polycrystalline structure obtains.

Pseudomorphic growth of the oxide layer up to a thickness of the order 100 Å has been suggested by Mott (1949). In many cases this would involve a strain greater than the yield strain of the bulk material; there is, however, evidence that material in thin layers can withstand stresses considerably greater than can the bulk material (Beams, Walker and Morton 1952). At thicknesses in excess of approximately 100 Å the strain must be released either by slip or by recrystallization. Mott shows theoretically the existence of a critical temperature below which growth to a limiting thickness occurs; after an induction period, during which recrystallization takes place, further oxidation occurs. There is, however, little reliable experimental evidence in favour of the existence of a strained layer of the thickness suggested. The growth of Cu_2O on copper (Mehl, McCandless and Rhines 1934) occurs with parallel orientation with a misfit of 18%. In the case of FeO on iron, the orientation observed is FeO (001) [110] ‖ Fe (001) [100]. Taking into account the lattice parameters, it is suggested that the formation of FeO from iron entails the expansion of the body-centred cell of α-iron to form a body-centred tetragonal lattice of axial ratio 1·414, in which the iron atoms lie in precisely the positions of the Fe atoms in (face-centred) FeO. In the case of aluminium, a face-centred form of alumina has been observed (Steinheil 1934) although this is not the normal structure of the oxide. Furthermore, in the growth of MgO on magnesium, a hexagonal form of the oxide is observed instead of the usual cubic form.

There is thus a general tendency for oxide to form in such a way that the positions occupied by the metal atoms in the oxide correspond as nearly as possible to those in the metal itself. If the misfit between the metal and oxide lattices is small, we may expect to observe extensive epitaxial growth up to large thicknesses of oxide with relief of strain by the formation of dislocations. In the event of a large misfit, the regions over which an ordered growth can occur would be limited and dislocations so frequent that a disordered, polycrystalline growth would be expected. These general features are exhibited in the oxides of iron, nickel and cobalt and are discussed in § 10.

§ 9. DISCUSSION OF RESULTS OF METALS ON ROCKSALT

We may now consider the extent to which the experimental results given in §§ 3–5 lend support to the various approaches to the problem of accounting for the phenomenon of epitaxy. We shall consider the orientations to be expected from the standpoints both of lattice misfit and of co-ordination.

9.1. *Epitaxy of Nickel* (f.c.c., $a = 3{\cdot}517$ Å)

For the (001) face of rocksalt, there are two orientations of the nickel which merit consideration. With the (001) plane of nickel parallel to (001) rocksalt we may have (1), [100] Ni ∥ [100] NaCl; (2), [100] Ni ∥ [110] NaCl. In the first orientation, the misfit between the Ni and NaCl lattices is $-37{\cdot}6\%$. The position of the Ni atom in the contact plane, relative to the chlorine ions of the substrate is, however, similar to that of the Ni atoms in nickel chloride. In each case, the Ni atoms are situated octahedrally relative to the chlorine sites. In the second orientation, the smaller misfit of $-11{\cdot}5\%$ obtains. In this case, however, the co-ordination number of some of the nickel sites is lower than for the parallel (1) orientation. Orientation (2) should be stable according to the van der Merwe model whereas orientation (1) should be unstable. From the co-ordination point of view, orientation (1) is the favourable one. This is the orientation which is observed, even in layers of average thickness about 10 Å, suggesting that the lattice strain, coupled with the effects of unfavourable co-ordination make orientation (2) unstable even in very thin layers.

Similar conditions apply to growth on the (110) plane of rocksalt. The nickel deposits with the (110) plane parallel to the rocksalt surface and with [001] Ni ∥ [001] NaCl. The rocksalt–nickel misfit is again $-37{\cdot}6\%$ whereas for the orientation [$\bar{1}$10] Ni ∥ [001] NaCl the smaller misfit of $-11{\cdot}5\%$ obtains but with unfavourable co-ordination conditions.

On the (111) plane of rocksalt, the nickel deposits with the (111) plane parallel to the surface. With this plane there is the possibility of growth with the small misfit of $+8{\cdot}0\%$ with [$\bar{1}$10] Ni ∥ [$\bar{2}$11] NaCl. When the second and third layers of nickel atoms are added in accordance with this orientation, however, the nickel atoms are not in the same sites relative to the chlorine ions as are the nickel atoms in $NiCl_2$. The correct co-ordination is obtained with parallel orientation $[\bar{1}10]$ Ni || $[\bar{1}10]$ NaCl in spite of the large misfit of −37·6%. Experimentally, both orientations are observed; the parallel orientation predominates and is the one which sustains as the film thickness is increased.

9.2. *Epitaxy of Cobalt* (f.c.c., $a = 3{\cdot}554$ Å)

The lattice constants of cobalt (3·55 Å) and nickel (3·52 Å) are close together; their chlorides are isomorphous. We may thus expect similar conditions to apply to the growth of cobalt as apply to nickel. For the most part similar behaviour is observed except for the case of very thin Co deposits on (001) rocksalt. Although a predominantly parallel orientation is observed, definite signs of the orientation [110] Co || [100] NaCl are obtained in thin layers. The misfit is slightly smaller (−10·7%) than that for the nickel case (−11·5%) although it would be surprising if this difference alone controlled the difference in behaviour. The non-parallel orientation is expected on the basis of misfit theory; the cobalt result suggests that such an orientation may be stable for a few atom layers but that it cannot persist to greater thicknesses.

9.3. *Epitaxy of Iron* (b.c.c., $a = 2{\cdot}861$ Å)

Ferrous chloride has a structure similar to that of nickel and cobalt chlorides. The unit cell of iron is, however, body-centred with a considerably smaller spacing of atoms along the [100] rows than is found in the Ni and Co lattices. On the (001) plane of rocksalt, iron deposits with its (001) plane parallel to the rocksalt. Considering the orientation [110] Fe || [100] NaCl, we may regard this as a face-centred tetragonal lattice in parallel orientation on the rocksalt. The Fe atoms in this orientation are in octahedral positions with respect to the Cl ions so that this orientation satisfies co-ordination requirements. The lattice misfit is −28·1%.

In the orientation [100] Fe || [100] NaCl, four unit cells of iron fit quite closely (misfit 1·7%) on one unit cell of rocksalt but with poor co-ordination for all the iron atoms in the contact plane. On the lattice misfit theory, this last-mentioned orientation should be stable whereas the first-mentioned should not. In fact, both orientations are observed in films formed at temperatures close to 450°c. The orientation satisfying co-ordination requirements predominates and is the orientation generally observed in films formed above 500°c. This suggests that the low-misfit orientation is relatively unstable and that recrystallization occurs at the higher temperature to give the orientation satisfying co-ordination conditions.

On the (110) plane of rocksalt, the following orientations are observed: (*a*), Fe(110)[001] || NaCl(110)[001]; (*b*), Fe(111)$[\bar{1}10]$ || NaCl(110)[001]. This case is of interest in so far as there is no orientation in which co-ordination requirements can be completely satisfied. Parallel orientation (*a*) with a one-to-one correspondence between cells gives a large misfit (−49%) and is one in which some of the Fe atoms are not in octahedral positions. Comparing four unit cells of Fe with one of NaCl for this orientation, a small misfit results (1·7%) but with still worse co-ordination. It should be noted that the orientation Fe(110)$[\bar{1}10]$ || NaCl(110)[001] does not satisfy co-ordination requirements. In fitting the (111) plane of iron on to the (110) plane of rocksalt, the orientation giving the nearest approach to satisfactory co-ordination is that with $[\bar{1}10]$ Fe $\|$ [001] NaCl, for which the misfit is −28%.

For the growth of Fe on the (111) plane of rocksalt, the orientation Fe(111)$[\bar{1}10]$ $\|$ NaCl(111)$[\bar{1}10]$ is one in which both lattice fit and co-ordination are admirably satisfied. This orientation is observed although it does not occur alone. A small amount of iron is found to be in the orientation with (111) planes parallel and with $[11\bar{2}]$Fe $\|$ $[\bar{1}10]$ NaCl although the misfit in this case is −12·1% and the distribution of atoms around the contact plane unsymmetrical. A strongly occurring orientation of Fe on (111) rocksalt is that with the (110) Fe as deposit plane. The orientation of the (110) planes is the one expected from a consideration of the symmetries of the planes (110) Fe and (111) NaCl. The atoms in the (110) Fe planes form isosceles triangles with base angles 54° 44′. With [001] Fe $\|$ $[\bar{1}10]$ NaCl there is the maximum correspondence between the sites of the two lattices.

The results discussed in this section are summarized in table 5.

Table 5

Metal	Orientation				Misfit (%)	Remarks
	Metal		Rocksalt			
Fe	(001)	[110]	(001)	[100]	−28·1	
(b.c.c.)	(110)	[001]	(110)	[001]	−49·0†	
	(111)	$[\bar{1}10]$	(110)	[001]	−28·1	
	(110)	[001]	(111)	$[\bar{1}10]$	−28·1	Along Fe [110].
	(111)	$[\bar{1}10]$	(111)	$[\bar{1}10]$	+ 1·7	
	(111)	$[\bar{2}11]$	(111)	$[\bar{1}10]$	−12·1	
Co	(001)	[100]	(001)	[100]	−36·8	
(f.c.c.)	(001)	[100]	(001)	[110]	−10·7	Very thin layers only.
	(110)	[001]	(110)	[001]	−36·8	
	(111)	$[\bar{1}10]$	(111)	$[\bar{1}10]$	−36·8	
Ni	(001)	[100]	(001)	[100]	−37·6	
(f.c.c.)	(110)	[001]	(110)	[001]	−37·6	
	(111)	$[\bar{1}10]$	(111)	$[\bar{1}10]$	−37·6	

† With four unit cells of Fe on one unit cell of NaCl, the misfit is +1·7%.

It is seen from the results for the growth of the metals on rocksalt that the orientations to be observed in all except the thinnest layers are generally those in which the correct co-ordination across the substrate–metal boundary is maintained. In very thin deposits, the orientation with lowest misfit is sometimes (but not always) observed. The position of minimum energy for the atoms in the first layer will clearly depend on whether or not there are other deposit layers above them. It is likely that the most stable positions for one or two monolayers may correspond to the lowest misfit orientation. When further layers are added, the orientation satisfying co-ordination becomes the stable form and is attained by recrystallization.

§ 10. THE GROWTH OF OXIDE FILMS ON FILMS OF NICKEL, COBALT AND IRON

The oxide films which have been studied are those occurring at 400–500°C and at a pressure 10^{-4} mm Hg and also those which form at room temperature and atmospheric pressure. The oxide film thicknesses are generally less than about 100 Å. For such films, oxidation probably occurs by diffusion of metal ions (Mott 1947) through the lattice, the growth law being logarithmic.

10.1. *Nickel and Cobalt Oxides*

(NiO: f.c.c., $a = 4{\cdot}172$ Å; CoO: f.c.c., $a = 4{\cdot}24$ Å)

The structures of NiO and CoO are similar to that of sodium chloride. Since the metals form face-centred cubic lattices, the relative positions of the metal atoms in the metals and oxide structures are similar. Co-ordination requirements are thus satisfied by parallel growth of the oxide on the metal and this form of growth is observed in thin oxide layers formed at temperatures above 450°c. The lattice misfit is considerable ($+18{\cdot}6\%$ for NiO; $+19{\cdot}5\%$ for CoO). The orientation [110] NiO || [100] Ni, which has the smaller misfit of -16% in the case of NiO, but for which the co-ordination conditions are not met, is not observed. The oxide films so formed show patterns characteristic of rotational slip, suggesting that they are under strain. As the oxide thickens, the single crystal form is lost and a polycrystalline structure results.

Oxidation at temperatures below about 300°c tends to produce an oriented oxide film with (111) NiO || (100) Ni. This form is also observed in films which are allowed to oxidize at room temperature for several hours. There are four possible orientations in which the oxide may lie with principal rows of atoms along the [110]-type Ni directions, viz. $[\bar{1}10]$ NiO || [110], $[\bar{1}\bar{1}0]$, $[\bar{1}10]$ and $[1\bar{1}0]$ Ni. The (111) plane of NiO is the plane of closest packing: there is frequently a tendency for crystalline deposits to form with the plane of closest packing parallel to the substrate plane. Similar behaviour is observed in the case of the system CoO–Co.

10.2. *The Oxidation of Iron Films*

(FeO: f.c.c., $a = 4{\cdot}294$ Å; Fe_3O_4: spinel, $a = 8{\cdot}42$ Å)

In the oxidation of iron films the first oxide layers formed are of FeO and the final oxide is the spinel Fe_3O_4. FeO may be considered as a close-packed cubic array of iron atoms in which the latter are in octahedral sites with respect to the oxygen atoms. The unit cell of the spinel structure contains 32 oxygen atoms in close-packed cubic array: iron atoms occupy 16 octahedral sites and eight tetrahedral sites. The spinel is thus formed by an ordering of vacancies in the octahedral sites and a filling of certain tetrahedral sites, the positions of the oxygen atoms being similar to those in FeO. The separation between oxygen atoms in the spinel is only very slightly greater ($<1\%$) than that in the FeO lattice. It is to be expected, therefore, that the spinel orientation will be governed directly by that of the FeO lattice. This is observed to be the case, the spinel showing parallel orientation to that of FeO.

On the (001) plane of iron, the oxide orientation observed is:

$$\text{FeO}\,(001)\,[100] \parallel \text{Fe}\,(001)\,[110].$$

The lattice misfit is $+6{\cdot}1\%$. The body-centred lattice of iron may be regarded as a face-centred tetragonal lattice with axes at 45° to those of the [100] iron directions. The observed orientation of the (face-centred) iron oxide is thus the one which preserves continuity of the face-centred arrangement of the iron atoms, suggesting that co-ordination requirements determine the orientation in this case. Parallel orientation ([100] FeO ‖ [100] Fe) is unsatisfactory, both from the co-ordination point of view and from that of lattice misfit (50·1%).

In the growth of oxide on (110) Fe, the orientation observed is that giving low misfit with FeO(110)$[\bar{1}10]$ ‖ Fe(110)[001]; the misfit is +6·1%. The orientation (111) FeO ‖ (110) Fe is also observed to a small extent, the cube face diagonal of the oxide lying along the cube edge of the iron. The misfits are 6·1% along the cube edge of Fe and 22·4% along the face diagonal. There is, however, a tendency among face-centred cubic materials to orientate with (111) planes parallel to the substrate, even when the substrate is amorphous. We see, therefore, that when this orientation occurs on a single-crystal substrate in which there is a close lattice fit in one direction only, a large misfit may be tolerated in other directions.

On the (111) planes of Fe, two oxide orientations are observed:

(*a*), FeO(111)$[\bar{1}10]$ ‖ Fe(111)$[\bar{1}10]$; (*b*), FeO(110)$[\bar{1}10]$ ‖ Fe(111)$[\bar{1}10]$.

Orientation (*a*) entails a lattice misfit of −24·9% but satisfies the requirement of a continuation of Fe structure across the boundary. The second orientation also entails a misfit of −24·9% but the relative positions of the iron atoms either side of the boundary are highly unsymmetrical. The degree of orientation observed in this case is much smaller than that obtained when the co-ordination requirements are satisfied.

§ 11. DISCUSSION OF THE OXIDATION RESULTS

The main features of the results obtained for the oxide layers on nickel, cobalt and iron at temperatures above 400°C are summarized in table 6. In the cases of nickel and cobalt, slow oxidation at room temperature results in growth with the (111) oxide plane parallel to the metal surfaces.

Table 6

Orientation					Misfit	Remarks
Metal			Oxide		(%)	
Fe	(001)	[100]	(001)	[110]	+ 6·1	Good fit and co-ordination. FeO lattice strained.
Fe	(110)	[001]	(110)	$[\bar{1}10]$	+ 6·1	Poor co-ordination.
Fe	(110)	[001]	(111)	$[\bar{1}10]$	+ 6·1 +22·4	Orientation not complete.
Fe	(111)	$[\bar{1}10]$	(110)	$[\bar{1}10]$	−24·9	Poor co-ordination.
Fe	(111)	$[\bar{1}10]$	(111)	$[\bar{1}10]$	−24·9	Co-ordination satisfied.
Ni :	Parallel growth on all planes				+18·6	Growth to few layers only.
Co :	Parallel growth on all planes				+19·5	Growth to few layers only.

The results obtained with nickel and cobalt films on the one hand and with iron films on the other give some support for the view of Frank and van der Merwe that a strained lattice may form provided a critical misfit of 9% (or 14% at low temperatures) is not exceeded. Direct evidence of a strained iron oxide lattice with a misfit of 6·1%, is found whereas nickel and cobalt oxides show the normal spacing. Although the difference in lattice spacings amounts to some 19% for these oxides, a limited amount of epitaxial growth occurs, degenerating into a polycrystalline form after a few molecular layers. Single-crystal growth of iron oxide persists to much greater thicknesses although there arises here the complication of the structure change with increasing thickness. The indication is that, whilst a strained lattice may form in cases where the misfit is small, epitaxial growth may occur with larger misfit than the critical value suggested by Frank and van der Merwe. Such growth is the more likely if there exists an orientation in which there is continuity of the metal structure across the boundary. The results given in §§ 9, 10 show that the misfit may be as large as 28%. Conversely, growth may occur where an orientation exists such that the lattice misfit is small even if co-ordination requirements are not fulfilled, as is illustrated by the parallel growth of FeO on the (110) plane of iron.

§ 12. CONCLUSIONS

When epitaxial growth occurs on a single-crystal substrate, either by deposition from the vapour or by oxidation, the orientation mainly observed is governed mainly by a tendency to continue the substrate structure across the boundary, so that the normal co-ordination of the atoms at the interface is preserved. The lattice misfit is of less importance than the co-ordination condition. However, if the lattice misfit is very large (~20–30%) incomplete orientation may occur or alternatively epitaxial growth may persist for only a small number of molecular layers. There is no indication that twinning and rotational slip are essential to the formation of an oriented overgrowth although they do serve to release growth strains in the deposit.

There is some evidence, from the oxidation results, for the formation of a strained lattice where the misfit is not large, as suggested by the theoretical treatment of Frank and van der Merwe. There are, however, many instances of epitaxy occurring when there is a larger misfit than that which would be expected to give a strained lattice. If a strained monolayer forms on the substrate surface and if the normal lattice spacing is attained in a very small number of molecular layers, the strained region may not be detected by the electron diffraction method. From the nature of the oxidation process, in which an initially complete metal lattice is changed to oxide by diffusion of atoms, it seems plausible that strained reigons may form to a greater extent than would be expected in a film forming by deposition from the vapour.

The existence of a minimum temperature for epitaxial growth to occur suggests that a certain degree of mobility in the atoms of the deposit is required in order that they may take up positions in an ordered array. For metals on rocksalt there is the further possibility that orientation may result from the existence of ionized atoms of the deposit metal.

REFERENCES

ACHARYA, H. K., 1948, *Ph.D. Thesis*, University of London.
BEAMS, J. W., WALKER, W. E., and MORTON, H. S., 1952, *Phys. Rev.*, **87**, 524.
BREDIG, G., and ALLOLIO, R., 1927, *Z. Phys. Chem.*, **126**, 52.
BREDIG, G., and SCHWARZ VON BERGKAMPF, E., 1931, *Z. Phys. Chem.* A, Bodenstein-Festband, 172.
BRÜCK, L., 1936, *Ann. Phys., Lpz.*, **26**, 233.
COCHRANE, W., 1936, *Proc. Phys. Soc.*, **48**, 723.
COLLINS, L. E., and HEAVENS, O. S., 1954, *Phil. Mag.*, **45**, 283.
DRABBLE, J. R., 1949, *Ph.D. Thesis*, University of London.
ENGEL, O. G., 1952, *J. Chem. Phys.*, **20**, 1174.
EVANS, D. M., and WILMAN, H., 1950, *Proc. Phys. Soc.* A, **63**, 298.

Finch, G. I., and Quarrell, A. G., 1933, *Proc. Roy. Soc.* A, **141**, 398; 1934, *Proc. Phys. Soc.*, **46**, 148.
Frank, F. C., and van der Merwe, J. H., 1949, *Proc. Roy. Soc.* A, **198**, 216.
Goche, O., and Wilman, H., 1939, *Proc. Phys. Soc.*, **51**, 625.
Haase, O., 1956, *Z. Naturforsch.*, **11a**, 46.
Heavens, O. S., 1952, *Proc. Phys. Soc.* B, **65**, 788; 1953, *Acta Cryst., Camb.*, **6**, 571.
Hull, A. W., 1917, *Phys. Rev.*, **10**, 661.
Ingersoll, L. R., and Hanawalt, J. D., 1929, *Phys. Rev.*, **34**, 972.
Juza, R., and Sachsze, W., 1943, *Z. anorg. Chem.*, **251**, 201.
Le Clerc, G., and Michel, A., 1939, *C. R. Acad. Sci., Paris*, **208**, 1583.
Lucas, L. N. D., 1951, *Proc. Phys. Soc.* A, **64**, 943.
Mehl, R. F., McCandless, E. L., and Rhines, F. N., 1934, *Nature, Lond.*, **134**, 1009.
Menzer, G., 1938 a, *Z. Kristallogr.*, **99**, 378; 1938 b, *Ibid.*, **99**, 410.
Mott, N. F., 1947, *Trans. Faraday Soc.*, **43**, 429; 1949, *Research*, **2**, 162.
Raether, H., 1950, *J. Phys. Radium*, **11**, 11.
Shirai, S., 1937, *Proc. Phys. Math. Soc. Japan*, **19**, 937; 1938, *Ibid.*, **20**, 855.
Smollett, M., and Blackman, M., 1951, *Proc. Phys. Soc.* A, **64**, 683.
Steinheil, A., 1934, *Ann. Phys., Lpz.*, **19**, 465.
Taylor, C. A., Hinde, R. M., and Lipson, H., 1951, *Acta. Cryst., Camb.*, **8**, 261.
Thomson, G. P., 1931, *Nature, Lond.*, **119**, 234.
Uyeda, R., 1942, *Proc. Phys. Math. Soc. Japan*, **24**, 809.
Wilman, H., 1940, *Proc. Phys. Soc.*, **52**, 195, 323; 1951, *Proc. Phys. Soc.* A, **64**, 329.
Wilson, A. J. C. 1942, *Proc. Roy. Soc.*, **180**, 277.

A Contribution to the Recombination Statistics of Excess Carriers in Semiconductors

By P. T. LANDSBERG
Marischal College, University of Aberdeen

MS. received 18*th September* 1956, *and in revised form* 7*th December* 1956

Abstract. The formalism of recombination statistics is extended in various directions. In the case of inter-band transitions, the conditions under which a recombination coefficient exists are investigated, and the effect of degeneracy of electron and hole gas is discussed. In the case of transitions in which traps participate, the formalism is extended to cover (*a*) degeneracy and (*b*) the case when impurity centres have several trapping levels. The more important special cases of the general formalism are considered. In all cases the possibility has been borne in mind that the departure from equilibrium may not be small.

---

§ 1. INTRODUCTION

THE purpose of this paper is to extend the statistical theory of lifetimes and recombination of excess current carriers in semiconductors. The present formalism is often based on a straightforward application of the mass action law. This gives useful results, but cannot be relied upon in the case of degeneracy or of energy dependent recombination coefficients. It also encounters difficulties if the departure from equilibrium can no longer be regarded as small, or if the condition of electrical neutrality fails to apply. A rather more broadly based theory appears, therefore, to be required. This should yield the results which have been found useful in the past. It should also yield new results in the fields just indicated, without going beyond its basic assumptions. An elementary theory which fulfills these requirements is given here; the assumptions will now be described.

First, it is supposed that even in a non-equilibrium stituation the conduction electrons, the holes in the valence band and the trapped electrons are all in equilibrium among themselves, even though these groups may not be in equilibrium with each other. This means that any non-equilibrium situation may be specified by three quasi-Fermi levels, which, when divided by $\boldsymbol{k}T$, will be denoted by F_e, F_h, F_t respectively. In equilibrium they are all equal to the equilibrium value, which will be denoted by F_0.

A second assumption is to suppose that the rate at which electrons are transferred to higher energy levels in a state of dynamic equilibrium is due to a factor, such as lattice vibrations, whose variation with electron and hole density can often be neglected. When discussing the recombination of holes and electrons by direct transitions between bands, for instance, one obtains the net recombination rate U_{eh} by subtracting from the actual recombination rate the electron–hole generation rate R_{eh}. The second assumption enables one to use for R_{eh} in steady non-equilibrium states the value which it has in the true equilibrium state, in which U_{eh} vanishes, and the actual recombination rate just cancels the generation rate.

Other assumptions made in this paper include the neglect of electron–electron interaction, and the hypothesis that the surfaces of constant energy in momentum space are spherical. Both these assumptions could be avoided, but the resulting complications in the formalism are not justified at present. Electrical neutrality is not always assumed.

§ 2. RECOMBINATION BY TRANSITIONS BETWEEN CONDUCTION AND VALENCE BANDS

The usual method of discussing this problem is to suppose that in a steady non-equilibrium state of the semiconductor, there is a recombination coefficient r, and an electron–hole generation rate g, such that the net rate of recombination per unit volume is

$$U = rnp - g = r(np - n_0 p_0). \qquad \ldots\ldots(2.1)$$

If $\delta n \equiv n - n_0$, $\delta p \equiv p - p_0$ be the excess concentrations of electrons and holes above their equilibrium values, the lifetimes of excess electrons and holes are then

$$\left.\begin{aligned} \tau_e &\equiv \delta n/U = r(cn_0 + p_0 + \delta p)^{-1}, \\ \tau_h &\equiv \delta p/U = r(n_0 + p_0/c + \delta n)^{-1}, \end{aligned}\right\} \qquad \ldots\ldots(2.2)$$

where $c \equiv \delta p/\delta n$. This is exact, and presupposes that the number of 'encounters' of electrons and holes in a fixed time interval, and in the steady state, is proportional to np and, further, that there is a probability r per encounter that recombination occurs. This approach was criticized in § 1. Its range of application will be assessed by means of the improved analysis specified in that section.

By (2.2), $\tau_e\,(\tau_h)$ is defined to be that time in which the loss of electrons (holes) in unit volume by the steady net recombination rate U is just equal to the steady concentration $\delta n\,(\delta p)$ of non-equilibrium electrons (holes). τ_e and τ_h are, therefore, average lifetimes for this steady state. These lifetimes are in general also of importance in connection with the return of the system to the equilibrium state after the perturbing influence has been removed. Though this type of transient situation will not be considered in the rest of this paper, a remark about it here may be of some use. Suppose U is $f(\delta n, x)\,\delta n$, where x stands for parameters, such as temperature and equilibrium Fermi level, which specify the physical state of the crystal, and $\delta n = \delta p$. Then $\tau = 1/f$ is the lifetime by (2.2). If the perturbing influence is switched off at time $t = 0$, one can assume $U = -d(\delta n)/dt$, so that $d(\delta n)/dt = -\delta n/\tau$. Thus τ determines the exponential decay of δn, which may be assumed to take place at any time t for a period which is small enough for τ to be practically a constant during it. More generally, it is found that τ governs the decay to equilibrium according to the relation

$$\delta n(t) = \delta n(0) \exp\left\{-\int_0^t \frac{dt}{\tau}\right\}.$$

In the special cases when τ is independent of time a simple exponential decay law is obtained. However, if $f(\delta n, x)$ is time-dependent it is not possible to define a generally valid lifetime by fitting a curve $\delta n \propto \exp(-t/\tau)$, though (2.2) gives still a possible definition in this case. This definition will therefore be used throughout this paper. If, for example, δn and δp are not negligible on the right-hand side of (2.2), the lifetimes τ become time-dependent for the decay process.

Let us write $F_0 + \delta F_e$ for F_e and $F_0 + \delta F_h$ for F_h. Let E_e, E_h be respectively the kinetic energy of an electron and a hole, and $\eta_e \equiv E_e/kT$, $\eta_h \equiv E_h/kT$. Let $A(\eta_e, \eta_h)$

be the number of electron–hole recombinations per unit volume per unit time, and for concentrations of one electron of energy E_e and one hole of energy E_h, each measured per unit volume. A appears therefore in units of cm^3 sec^{-1}. Let $N_e(\eta_e)$, $N_h(\eta_h)$ be the number of quantum states per unit energy range per unit volume. The recombination rate (in cm^{-3} sec^{-1}) is then given by

$$U_{eh} = (\boldsymbol{k}T)^2 \int_{\eta_c}^{\infty} \int_{-\infty}^{\eta_v} N_e N_h A \left[P_e P_h - P_{e0} P_{h0}\right] d\eta_h \, d\eta_e, \qquad \ldots\ldots(2.3)$$

where the occupation probabilities are

$$P_e \equiv [1 + \exp(\eta_e - F_0 - \delta F_e)]^{-1}, \quad P_h \equiv [1 + \exp(F_0 + \delta F_h - \eta_h)]^{-1}. \qquad \ldots\ldots(2.4)$$

P_{e0}, P_{h0} are the corresponding equilibrium values. η_c, η_v are respectively the energies, divided by $\boldsymbol{k}T$, at the bottom of the conduction band and at the top of the valence band.

The electron–hole pair generation rate is

$$R_{eh} \equiv (\boldsymbol{k}T)^2 \int_{\eta_c}^{\infty} N_e P_{e0} \, d\eta_e \int_{-\infty}^{\eta_v} A N_h P_{h0} \, d\eta_h. \qquad \ldots\ldots(2.5)$$

Hence,

$$R_{eh} = \bar{A} n_0 p_0. \qquad \ldots\ldots(2.6)$$

where $\bar{A}$ is the probability $A(\eta_e, \eta_h)$, as averaged over the electrons in the conduction band and the holes in the valence band, in the true equilibrium state.

Defining now

$$B_e \equiv \frac{(\boldsymbol{k}T)^2}{\delta F_e} \int_{\eta_c}^{\infty} N_e (P_e - P_{e0}) \, d\eta_e \int_{-\infty}^{\eta_v} A P_{h0} N_h \, d\eta_h, \qquad \ldots\ldots(2.7\,a)$$

$$B_h \equiv \frac{(\boldsymbol{k}T)^2}{\delta F_h} \int_{\eta_c}^{\infty} N_e P_{e0} \, d\eta_e \int_{-\infty}^{\eta_v} A (P_h - P_{h0}) N_h \, d\eta_h, \qquad \ldots\ldots(2.7\,b)$$

$$B_{eh} \equiv \frac{(\boldsymbol{k}T)^2}{\delta F_e \delta F_h} \int_{\eta_c}^{\infty} N_e (P_e - P_{e0}) \, d\eta_e \int_{-\infty}^{\eta_v} A (P_h - P_{h0}) N_h \, d\eta_h, \qquad \ldots\ldots(2.7\,c)$$

it is found that

$$U_{eh} = B_e \delta F_e + B_h \delta F_h + B_{eh} \delta F_e \delta F_h. \qquad \ldots\ldots(2.8)$$

Also

$$\delta_n = a_e \delta F_e, \quad \delta_p = a_h \delta F_h, \qquad \ldots\ldots(2.9)$$

where

$$a_e \equiv \frac{\boldsymbol{k}T}{\delta F_e} \int_{\eta_c}^{\infty} N_e (P_e - P_{e0}) \, d\eta_e, \quad a_h \equiv \frac{\boldsymbol{k}T}{\delta F_h} \int_{-\infty}^{\eta_v} N_h (P_h - P_{h0}) \, d\eta_h. \qquad \ldots\ldots(2.10)$$

It follows that the lifetimes of excess electrons and excess holes for band-to-band transitions are formally given by

$$1/\tau_e = \sum_{j=1}^{3} (1/\tau_{ej}), \quad 1/\tau_h = \sum_{j=1}^{3} (1/\tau_{hj}); \qquad \ldots\ldots(2.11\,a)$$

here

$$\tau_{e1} \equiv \frac{a_e}{B_e}, \quad \tau_{e2} \equiv \frac{a_h \delta n}{B_h \delta p}, \quad \tau_{e3} \equiv \frac{a_e a_h}{B_{eh} \delta p}, \qquad \ldots\ldots(2.11\,b)$$

$$\tau_{h1} \equiv \frac{a_h}{B_h}, \quad \tau_{h2} \equiv \frac{a_e \delta p}{B_e \delta n}, \quad \tau_{h3} \equiv \frac{a_e a_h}{B_{eh} \delta n}. \qquad \ldots\ldots(2.11\,c)$$

The component lifetimes have simple interpretations: τ_{e1} is due to the departure from equilibrium of the electron concentration only, τ_{e2} due to the departure from equilibrium of the hole concentration only, and τ_{e3} is a cross term which is of the second order of smallness if all deviations from equilibrium are small. Similar interpretations apply to the lifetimes of excess holes.

By expanding P_e and P_h about their equilibrium values, one can, for small disturbances, express the net recombination in terms of the pair generation rate R_{eh}:

$$U_{eh} = R_{eh}\left[\frac{\mu_e}{a_e}\delta n + \frac{\mu_h}{a_h}\delta p\right]. \qquad \ldots\ldots(2.12)$$

The μ's are given by

$$\left.\begin{aligned} \mu_e R_{eh} &\equiv (\boldsymbol{k}T)^2 \iint N_e N_h A\, P_{h0} (\partial P_e/\partial F_e)_0\, d\eta_e\, d\eta_h, \\ \mu_h R_{eh} &\equiv (\boldsymbol{k}T)^2 \iint N_e N_h A\, P_{e0} (\partial P_h/\partial F_h)_0\, d\eta_e\, d\eta_h. \end{aligned}\right\} \qquad \ldots\ldots(2.13)$$

It will be found later (e.g. in equations (5.4) and (5.15)) that (2.12) is a useful way of writing the result.

In view of (2.12) it is desirable to consider R_{eh} as a function of the reduced Fermi level F_0 as this increases from valence to conduction band. If the hole gas is degenerate the electron gas is non-degenerate, and R_{eh} rises approximately as $[\exp F_0] \int A N_h\, d\eta_h$. When both gases are non-degenerate R_{eh} has reached a constant plateau, and it falls off again approximately as $[\exp(-F_0)] \int A N_e\, d\eta_e$ as the electron gas becomes degenerate. The value, R' say, of R_{eh} at this constant plateau can be used to evaluate the average

$$\begin{aligned} \bar{A} &= \tfrac{1}{4} R' \left[h^2/2\pi \boldsymbol{k} T \sqrt{(m_e m_h)}\right]^3 \exp \eta_g = R'/n_i^2 \\ &= 4{\cdot}29\, R' T^{-3} (m^2/m_e m_h)^{3/2} \exp \eta_g \times 10^{-32}\ \mathrm{cm^3\,sec^{-1}}, \end{aligned} \qquad \ldots\ldots(2.14)$$

where η_g is the energy gap E_g divided by $\boldsymbol{k}T$, and n_i^2 is the product of n_0 and p_0 in the classical limit. This is in agreement with what is known of the pair generation rate in indium antimonide. A plateau in R_{eh}, and a drop in R_{eh}, as the electron gas becomes degenerate, have both been found empirically in this case (Mackintosh and Allen 1955, figure 4). Equation (2.14) is used to evaluate $\bar{A}$ in the table. It is assumed in this table that photon-radiative equilibrium and detailed balance both hold and that the effect of traps may be neglected.

Mean Transition Probabilities $\bar{A}$

Substance	R′ (10^{21} cm³ sec⁻¹)†	n_i (10^{16} cm⁻³)	$\bar{A}$ (10^{-11} cm³ sec⁻¹)
PbSe	20	2·3	3·8
PbTe	8·6	1·3	5·1
PbS	0·048	0·35	0·39
InSb	10	1·56	4·1

† Evaluated by integration from optical data by Mackintosh (1956 a) and Mackintosh and Allen (1955).

If R_{eh} were the only factor which varies with composition in (2.12), the lifetime would have a minimum for samples having both electron and hole gases non-degenerate. In fact, however, the second factor in (2.12) tends to dominate, and to act in the opposite sense to R_{eh} (§ 5).

§ 3. Trapping at Localized Levels

This section develops the work of Shockley and Read (1952), by extending their theory to defects which have several trapping levels, while at the same time allowing an arbitrary degree of degeneracy, and catering for departures from equilibrium carrier concentrations which may also be arbitrary within a wide range. An assumption to be made in this section is to suppose that direct recombination by transitions between conduction and valence bands can be neglected.

The probability that the *i*th *j*-electron state of a centre is occupied is given by (Landsberg 1956)

$$P_{ij}=\frac{\exp\left[jF_t-\eta_{ij}\right)}{\sum_{k=0}^{M} Z_k \exp(kF_t)}, \qquad \ldots\ldots(3.1)$$

where η_{ij} is $(1/\boldsymbol{k}T)$ times the energy associated with the *i*th *j*-electron state of a centre, Z_k is the ordinary partition function $\Sigma_r \exp(-\eta_{rk})$ of a centre which has trapped k electrons, and M is the number of electrons which a centre can trap. By way of example, an estimate of Z_1 in the presence of a magnetic field is given in Appendix I. If G be the number of centres present, then the rate of capture of conduction band electrons of the energy range (E_e, E_e+dE_e) by the *i*th *j*-electron state of the centres is

$$P_e(E_e)\,N_e(E_e)\,A(E_e\,;i,j)\,G\,P_{ij-1}\,dE_e \qquad (j=1,2,\cdots M).$$

It is to be noted that the mean number of centres which can capture an electron so that it will go into the *i*th *j*-electron state is equal to the mean number of centres $G\,P_{ij-1}$ which have their *i*th $(j-1)$-electron state occupied†. Similarly the rate of emission of electrons from the *i*th *j*-electron state of the centres into the energy range (E_e, E_e+dE_e) of the conduction band is

$$G\,P_{ij}\,A(i,j\,;E_e)(1-P_e)\,N_e\,dE_e \qquad (j=1,2,\ldots M).$$

In these expressions A is a transition probability of the type discussed in § 2. The net capture rate is

$$dU(E_e\,;i,j)=P_e\,N_e\,A(E_e\,;i,j)\,G\,P_{ij-1}\left[1-\frac{P_{ij}}{P_{ij-1}}\frac{A(i,j\,;\,E_e)}{A(E_e\,;\,i,j)}\frac{1-P_e}{P_e}\right]dE_e.$$

The ratio of the A's can be identified from the thermal equilibrium case, whence

$$dU(E_e\,;i,j)=P_eN_e\,A(E_e\,;i,j)\,G\,\frac{\exp\left[(j-1)F_t-\eta_{ij-1}\right]}{\sum_{k=0}^{M} Z_k \exp(kF_t)}\left[1-\exp(F_t-F_e\right]dE_e. \qquad \ldots\ldots(3.2)$$

Similarly the net capture rate of holes by traps is found to be

$$dU(E_h\,;i,j)=P_hN_h\,A(E_h\,;i,j)\,G\,\frac{\exp(jF_t-\eta_{ij})}{\sum_{k=0}^{M} Z_k \exp(kF_t)}\left[1-\exp(F_h-F_t)\right]dE_h. \qquad \ldots\ldots(3.3)$$

Integrating (3.2) and (3.3) over energy, and summing over i and j, one finds for the total capture rates

$$U_{et}=\frac{G[1-\exp(F_t-F_e)]}{\sum_{k=0}^{M} Z_k \exp(kF_t)}\sum_{r=1}^{M} c_{r-1}\exp\left[(r-1)\,F_t\right], \qquad \ldots\ldots(3.4)$$

† See Appendix II

$$U_{\mathrm{ht}} = \frac{G[1-\exp(F_{\mathrm{h}}-F_{\mathrm{t}})]}{\sum\limits_{k=0}^{M} Z_k \exp(kF_{\mathrm{t}})} \sum_{r=1}^{M} d_r \exp(rF_{\mathrm{t}}). \qquad \ldots\ldots(3.5)$$

Here the c's and d's have the meanings

$$\left.\begin{aligned} c_{j-1} &\equiv \sum_i \exp(-\eta_{i\,j-1}) \int_{E_{\mathrm{e}}}^{\infty} P_{\mathrm{e}} N_{\mathrm{e}} A(E_{\mathrm{e}}; i,j)\, dE_{\mathrm{e}}, \\ d_j &\equiv \sum_i \exp(-\eta_{ij}) \int_{-\infty}^{E_{\mathrm{v}}} P_{\mathrm{h}} N_{\mathrm{h}} A(E_{\mathrm{h}}; i,j)\, dE_{\mathrm{h}}, \end{aligned}\right\} \qquad \ldots\ldots(3.6)$$

where $j=1, 2, \ldots M$ and $\eta_{i0} \equiv 0$. In a steady state $U_{\mathrm{et}} = U_{\mathrm{ht}}$ and this condition yields an equation for F_{t} :

$$\sum_{j=1}^{M} \{c_{j-1}[1-\exp(F_{\mathrm{t}}-F_{\mathrm{e}})] + d_j[\exp F_{\mathrm{h}} - \exp F_{\mathrm{t}}]\} \exp[(j-1)F_{\mathrm{t}}] = 0. \qquad \ldots\ldots(3.7)$$

This equation is of order M in $\exp(F_{\mathrm{t}})$. Its solutions are

$$\exp F_{\mathrm{t}} = \frac{c_0 + d_1 \exp F_{\mathrm{h}}}{c_0 \exp(-F_{\mathrm{e}}) + d_1} \text{ if } \quad M=1; \qquad \ldots\ldots(3.8)$$

and, if $M=2$,

$$\exp F_{\mathrm{t}} = \frac{|\theta|}{2[c_1 \exp(-F_{\mathrm{e}}) + d_2]} \left\{ \left[1 + \frac{4[c_1 \exp(-F_{\mathrm{e}}) + d_2][c_0 + d_1 \exp F_{\mathrm{h}}]}{\theta^2} \right]^{1/2} \pm 1 \right\}. \qquad \ldots\ldots(3.9)$$

Here $\theta \equiv c_1 + d_2 \exp(F_{\mathrm{h}}) - c_0 \exp(-F_{\mathrm{e}}) - d_1$ and the top sign must be used if $\theta > 0$, the bottom sign if $\theta < 0$. Substitution of the solution of (3.7) for given M into (3.4) or (3.5) yields the steady state recombination rate U, and hence the lifetimes may be found. This completes the formal solution of the problem.

The steady state recombination rate is in the case $M=1$

$$U_{\mathrm{et}} = \frac{c_0 d_1 G[1-\exp(F_{\mathrm{h}}-F_{\mathrm{e}})]}{c_0[Z_0 \exp(-F_{\mathrm{e}}) + Z_1] + d_1[Z_0 + Z_1 \exp F_{\mathrm{h}}]}. \qquad \ldots\ldots(3.10)$$

Eor this case, consider two specimens of a semiconductor which are identical in all respects, except that the defects in the second specimen have certain additional trapping states. Symbols referring to this second specimen will be given primes, Hence

$$Z_1' > Z_1, \quad c_0' > c_0, \quad d_1' > d_1. \qquad \ldots\ldots(3.11)$$

To ensure that the degree of ionization of the defects in the steady non-equilibrium state under consideration is the same in both specimens, $\Sigma_i P_{i1} = \Sigma_i P_{i1}'$, using the notation of (3.1). This leads to

$$0 < \exp(F_{\mathrm{t}}' - F_{\mathrm{t}}) = Z_1/Z_1' < 1. \qquad \ldots\ldots(3.12)$$

Thus, as more trapping levels are introduced into the second specimen, its buasi-Fermi level for traps is lowered in order to keep $\Sigma_i P_{i1}'$ constant. F_{e} and F_{h} areassumed the same in the two specimens. Using the integrated form of (3.2) with $M=1$,

$$\begin{aligned} U_{\mathrm{et}}' - U_{\mathrm{et}} = \frac{G}{Z_0 + Z_1 \exp F_{\mathrm{t}}} \{ & [c_0' - c_0][1 - \exp(F_{\mathrm{t}}' - F_{\mathrm{e}})] \\ & + c_0 \exp(-F_{\mathrm{e}})[\exp F_{\mathrm{t}} - \exp(F_{\mathrm{t}}')] \}. \end{aligned} \qquad \ldots\ldots(3.13)$$

It follows from (3.4), (3.11) and (3.12) that

$$\text{if } U_{\mathrm{et}}' > 0 \text{ then } U_{\mathrm{et}}' > U_{\mathrm{et}} > 0. \qquad \ldots\ldots(3.14)$$

Thus, if the net capture rate of excess conduction electrons by traps is positive for both specimens, then this rate is increased by the provision of additional trapping levels. This is in agreement with the simple consideration that if a constant probability $\Sigma_i P_{i1}$ of a defect having captured an electron is spread over a greater number of levels, the number of vacant states at which electrons may be trapped is increased, and hence the net recombination rate is increased.

For the valence band, using (3.5),

$$U_{\mathrm{ht}} - U_{\mathrm{ht}}' = \frac{G}{Z_0 + Z_1 \exp F_{\mathrm{t}}} \{[d_1' - d_1][\exp F_{\mathrm{h}} - \exp F_{\mathrm{t}}'] + d_1[\exp F_{\mathrm{t}} - \exp F_{\mathrm{t}}']\}. \qquad \ldots\ldots(3.15)$$

By (3.5), (3.11) and (3.12),

$$\text{if } U_{\mathrm{ht}} > 0, \quad \text{then} \quad U_{\mathrm{ht}} > U_{\mathrm{ht}}' > 0. \qquad \ldots\ldots(3.16)$$

Thus, if the net capture rate of holes by traps is positive for both specimens, then this rate is decreased if the number of trapping levels per defect is increased.

The suffix 0 will now be used to label the true equilibrium values of quantities which have also been defined for non-equilibrium situations. With this notation, define, in analogy with (2.7),

$$\gamma_{j-1}\delta F_{\mathrm{e}} \equiv c_{j-1} - (c_{j-1})_0, \quad \delta_j \delta F_{\mathrm{h}} \equiv d_j - (d_j)_0 \quad (j = 1, 2, \ldots, M). \qquad \ldots\ldots(3.17)$$

Using (2.9) and (3.17), it is possible to re-express the net recombination rate (3.10) in terms of quantities which specify the equilibrium state of the system and those which specify the departure from equilibrium (quantities which are supposed to be known). Hence

$$\tau_{\mathrm{e}} = \frac{a_{\mathrm{e}}\delta F_{\mathrm{e}} P}{G(\delta F_{\mathrm{e}} - \delta F_{\mathrm{h}})(d_1)_0} [Z_0 \exp(-F_0 - \delta F_{\mathrm{e}}) + Z_1] \left\{1 + \left(\frac{d_1}{c_0}\right)_0 Q \exp(F_0 + \delta F_{\mathrm{e}})\right\}, \qquad \ldots\ldots(3..18)$$

$$\tau_{\mathrm{h}} = (a_{\mathrm{h}}\delta F_{\mathrm{h}}/a_{\mathrm{e}}\delta F_{\mathrm{e}})\tau_{\mathrm{e}}, \qquad \ldots\ldots(3.19)$$

where

$$\left.\begin{aligned} P &\equiv \frac{\delta F_{\mathrm{e}} - \delta F_{\mathrm{h}}}{1 - \exp(\delta F_{\mathrm{h}} - \delta F_{\mathrm{e}})} \frac{(d_1)_0}{(d_1)_0 + \delta_1 \delta F_{\mathrm{h}}} \\ Q &\equiv \frac{(c_0)_0[(d_1)_0 + \delta_1\delta F_{\mathrm{h}}]}{(d_1)_0[(c_0)_0 + \gamma_0 \delta F_{\mathrm{e}}]} \frac{Z_0 + Z_1 \exp(F_0 + \delta F_{\mathrm{h}})}{Z_0 + Z_1 \exp(F_0 + \delta F_{\mathrm{e}})}. \end{aligned}\right\} \qquad \ldots\ldots(3.20)$$

In the limit of very small disturbances, $P = Q = 1$ and (3.18) can be shown to go over into equation (5.5) of Shockley and Read (1952), provided certain additional assumptions are made.

It is instructive to derive alternative expressions for the lifetimes in terms of the actual equilibrium emission rates of electrons and holes by traps:

$$R_{\mathrm{et}} = \frac{G}{\sum_{k=0}^{M} Z_k \exp(kF_0)} \sum_{j=1}^{M} (c_{j-1})_0 \exp[(j-1)F_0] \rightarrow \frac{(c_0)_0 G}{Z_0 + Z_1 \exp F_0}, \qquad \ldots\ldots(3.21)$$

$$R_{\mathrm{ht}} = \frac{G}{\sum_{k=0}^{M} Z_k \exp(kF_0)} \sum_{j=1}^{M} (d_j)_0 \exp(jF_0) \rightarrow \frac{(d_1)_0 G}{Z_0 \exp(-F_0) + Z_1}. \qquad \ldots\ldots(3.22)$$

The last expressions apply to the case $M=1$. Equation (3.18) becomes

$$\tau_e = \frac{a_e \delta F_e}{\delta F_e - \delta F_h}\left\{\frac{S}{R_{et}} + \frac{T}{R_{ht}}\right\}, \qquad \ldots\ldots(3.23)$$

where

$$\left.\begin{aligned} S &\equiv \frac{\delta F_e - \delta F_h}{1-\exp(\delta F_h - \delta F_e)}\,\frac{(c_0)_0}{(c_0)_0+\gamma_0 \delta F_e}\,\frac{Z_0 + Z_1 \exp(F_0 + \delta F_h)}{Z_0 + Z_1 \exp F_0}, \\ &\text{and} \\ T &\equiv \frac{\delta F_e - \delta F_h}{1-\exp(\delta F_h - \delta F_e)}\,\frac{(d_1)_0}{(d_1)_0+\delta_1 \delta F_h}\,\frac{Z_0 \exp(-\delta F_e) + Z_1 \exp F_0}{Z_0 + Z_1 \exp F_0}. \end{aligned}\right\} \qquad \ldots\ldots(3.24)$$

$S=T=1$ in the limit of small disturbances. Equations (3.19) and (3.23) give the analogous expression for τ_h.

Equation (3.23), like (2.12), shows the close relationship which exists between the non-equilibrium rates U and the equilibrium rates R. The lifetime of excess electrons due to trapping may therefore be regarded as a sum of two component lifetimes; one of them is governed by the equilibrium emission (or capture) rate of electrons by traps, the other by the equilibrium emission (or capture) rate of holes by traps. By (3.21) and (3.22) for $M=1$

$$\frac{R_{et}}{R_{ht}} = \frac{\sum_i \int_{\eta_c}^{\infty} P_{e0} N_e A \, d\eta_e}{\sum_i \exp(F_0 - \eta_{i1}) \int_{-\infty}^{\eta_v} P_{h0} N_h A \, d\eta_h}. \qquad \ldots\ldots(3.25)$$

It is seen from (3.25) that if one imagines the Fermi level to be raised isothermally by appropriate changes in composition, R_{et}/R_{ht} increases.

If $\delta F_e = \delta F_h$ the recombination rates vanish and the lifetimes τ_e, τ_h become infinitely long. For, by (3.8),

$$\exp(\delta F_t - \delta F_e) = \frac{c_0 + d_1 \exp(F_0 + \delta F_h)}{c_0 + d_1 \exp(F_0 + \delta F_e)}. \qquad \ldots\ldots(3.26)$$

It follows that if any two of the quantities δF_e, δF_h, δF_t are equal, then they are all three equal, and equilibrium holds once again. A vanishing recombination rate is therefore expected.

The condition of electrical neutrality for the cases considered in this section takes the form (cf. equation (3.1))

$$\delta n + G \sum_{j=1}^{M} j\,\delta P_j = \delta p, \quad P_j \equiv \sum_i P_{ij}, \qquad \ldots\ldots(3.27)$$

in generalization of equation (A 4) of Shockley and Read. P_j is the probability of finding a centre which has trapped j electrons. From (3.1) and (2.9), (3.27) becomes

$$a_e \delta F_e + G t \, \delta F_t = a_h \delta F_h. \qquad \ldots\ldots(3.28)$$

t is here the mean square deviation of the number of electrons trapped at the impurities, as may be seen from

$$t \equiv \sum_{j=1}^{M} j^2 P_j - \left(\sum_{k=1}^{M} k P_k\right)^2 = \sum_{j=1}^{M}\left[j - \sum_k k P_k\right]^2 P_j. \qquad \ldots\ldots(3.29)$$

If a centre can trap only one electron, then $t = P_1(1-P_1)$. Equation (3.28), together with the condition (3.7) for a steady state, determines any one of the three quantities δF_e, δF_h, δF_t in terms of the other two. If $M=1$, eliminating F_t, it is found that

$$\left[1+\frac{a_e}{GP_1(1-P_1)}\right]\delta F_e-\ln\left[c_0+d_1\exp(F_0+\delta F_e)\right]=\frac{a_h\delta F_h}{GP_1(1-P_1)} -\ln\left[c_0+d_1\exp(F_0+\delta F_h)\right]. \qquad \ldots\ldots(3.30)$$

This equation simplifies in certain simple cases (cf. § 5). Equation (3.28) shows that the case $\delta F_e=\delta F_h=\delta F_t$, envisaged in the preceding paragraph, is ruled out by the condition of electrical neutrality, except for the trivial case $\delta F_e=\delta F_h=\delta F_t=0$.

§ 4. DISCUSSION OF INTEGRALS

Definitions in terms of integrals have been given for

$$a_e, a_h \,(\text{eqn } (2.10)); \quad B_e, B_h, B_{eh} \,(\text{eqn } (2.7)); \quad \gamma_{j-1}, \delta_j \,(\text{eqn } (3.17)) \qquad \ldots\ldots(4.1)$$

and for

$$R_{eh}, \bar{A} \,(\text{eqn } (2.5)); \quad c_j, d_j \,(\text{eqn } (3.6)). \qquad \ldots\ldots(4.2)$$

The quantities (4.2) are closely related to averages of the probabilities A. The symbols (4.1) have more complicated definitions. They involve integrals of the form

$$\left.\begin{aligned} I &\equiv \frac{kT}{\delta F_e}\int_{\eta_c}^{\infty}(P_e-P_{e0})f(E_e,x)\,d\eta_e, \text{ where } f(E_c,x)=0, \\ \text{and} \quad J &\equiv \frac{kT}{\delta F_h}\int_{-\infty}^{\eta_v}(P_h-P_{h0})g(E_h,x)\,d\eta_h, \text{ where } g(E_v,x)=0. \end{aligned}\right\} \qquad \ldots\ldots(4.3)$$

x stands here for variables other than the variable of integration. By partial integration,

$$I=kT\frac{1-\exp(-\delta F_e)}{\delta F_e}\int_{\eta_c}^{\infty}P_{e0}\frac{d}{d\eta_e}\left[\frac{1+\exp(\eta_e-F_0)}{1+\exp(\eta_e-F_0-\delta F_e)}f(E_e,x)\right]d\eta_e, \qquad \ldots\ldots(4.4\,a)$$

$$J=kT\frac{\exp(\delta F_h)-1}{\delta F_h}\int_{-\infty}^{\eta_v}P_{h0}\frac{d}{d\eta_h}\left[\frac{1+\exp(F_0-\eta_h)}{1+\exp(F_0+\delta F_h-\eta_h)}g(E_h,x)\right]d\eta_h. \qquad \ldots\ldots(4.4\,b)$$

Suppose the first factor in the square brackets of (4.4) can be approximated by a constant (denoted by λ in the case of (4.4 *a*), and μ in the case of (4.4 *b*)). Suppose also that the second factors are of the form

$$f(E_e,x)=f_0(E_e-E_c)^s, \quad g(E_h,x)=g_0(E_v-E_h)^t; \quad s,\ t>0, \qquad \ldots\ldots(4.5)$$

where f_0, g_0 are constants. Hence, using the Fermi integrals

$$I(r,a)\equiv\frac{1}{\Gamma(r+1)}\int_0^{\infty}\frac{x^r\,dx}{1+\exp(x-a)}, \quad r>0, \qquad \ldots\ldots(4.6)$$

it is found that

$$I=\frac{1-\exp(-\delta F_e)}{\delta F_e}\lambda f_0(kT)^{s+1}\,\Gamma(s+1)\,I(s-1,F_0-\eta_c), \qquad \ldots\ldots(4.7\,a)$$

$$J=\frac{\exp(\delta F_h)-1}{\delta F_h}\mu g_0(kT)^{t+1}\,\Gamma(t+1)\,I(t-1,\eta_v-F_0). \qquad \ldots\ldots(4.7\,b)$$

These formulae can be applied to the special case when the electron and hole gases are non-degenerate, but the disturbances are otherwise arbitrary. In this case

$$\lambda \simeq \exp(\delta F_e), \quad \mu \simeq \exp(-\delta F_h), \quad I(r,a) \simeq \exp a. \qquad \ldots\ldots(4.8)$$

It may also be applied to the case when the departure from equilibrium is small, but the degree of degeneracy is arbitrary, when

$$\lambda \simeq 1, \quad \mu \simeq 1. \qquad \ldots\ldots(4.9)$$

Another possibility is that the electron gas or the hole gas is degenerate. In this case one finds from (4.4) by a partial integration, and using well known formulae for integrations involving the Fermi–Dirac function,

$$I \simeq kT \frac{1-\exp(-\delta F_e)}{\delta F_e} \frac{2}{1+\exp(-\delta F_e)} f(F_0, x), \qquad \ldots\ldots(4.10\,a)$$

$$J \simeq kT \frac{\exp(\delta F_h)-1}{\delta F_h} \frac{2}{1+\exp(\delta F_h)} g(F_0, x). \qquad \ldots\ldots(4.10\,b)$$

§ 5. SPECIAL CASES

5.1. *Electron and Hole Gas Non-Degenerate*

(a) Inter-band transitions: δn, δp arbitrary.

In this case the probabilities (2.4) can be replaced by simple exponentials. Hence, by (2.5), (2.7), and (2.10)

$$R_{eh} = (kT)^2 \int_{\eta_c}^{\infty} N_e \exp(-\eta_e)\, d\eta_e \int_{-\infty}^{\eta_v} A N_h \exp \eta_h\, d\eta_h. \qquad \ldots\ldots(5.1)$$

$$B_e = \frac{\exp(\delta F_e)-1}{\delta F_e} R_{eh}, \qquad \ldots\ldots(5.2\,a)$$

$$B_h = \frac{\exp(-\delta F_h)-1}{\delta F_h} R_{eh}, \quad B_{eh} = \frac{[\exp \delta F_e - 1]\,[\exp(-\delta F_h)-1]}{\delta F_e \delta F_h} R_{eh}. \qquad \ldots\ldots(5.2\,b, c)$$

$$a_e = \frac{\exp \delta F_e - 1}{\delta F_e} \eta_0, \qquad \ldots\ldots(5.3\,a)$$

$$a_h = \frac{\exp(-\delta F_h)-1}{\delta F_h} p_0. \qquad \ldots\ldots(5.3\,b)$$

Substituting (5.2) into (2.8), and using (2.6), the recombination rate is

$$U_{eh} = R_{eh}[\exp(\delta F_e - \delta F_h) - 1]\, n = \bar{A}(np - n_0 p_0), \qquad \ldots\ldots(5.4), (5.5)$$

so that the simple results (2.1) and (2.2) are regained.

One can conclude that a recombination coefficient r exists in this case, and is given by the average transition probability $\bar{A}$. $\bar{A}$ is independent of the Fermi level and therefore of the carrier concentrations n_0, p_0, n, p. Thus the general theory yields the usual results for inter-band transitions (Shockley and van Roosbroeck 1954).

The recombination rate U vanishes if $\delta F_e = \delta F_h$. The explanation is similar to that given at the end of § 3. In the case of charge neutrality, $\delta F_e = \delta F_h$ implies again $\delta F_e = \delta F_h = 0$.

(*b*) *Trapping: δn and δp restricted.*

In the case of non-degenerate electron and hole gases, the lifetime (3.23) simplifies if the assumption of electrical neutrality is made ($\delta n = \delta p$). This implies that the degree of ionization of the traps is not appreciably altered by the disturbance. One finds

$$\tau_e = \frac{[\exp(\delta F_e) - 1][\exp(-\delta F_h) - 1] n_0 p_0}{[\exp(-\delta F_h) - 1]\delta F_e p_0 - [\exp \delta F_e - 1]\delta F_h n_0}\left[\frac{S}{R_{et}} + \frac{T}{R_{ht}}\right]. \quad \ldots\ldots(5.6)$$

If the departure from equilibrium is small, this becomes

$$\tau_e = \frac{n_0 p_0}{n_0 + p_0}\left[\frac{1}{R_{et}} + \frac{1}{R_{ht}}\right], \quad \ldots\ldots(5.7)$$

and generalizes equation (5.5) of Shockley and Read. The lifetime (5.7) depends on the Fermi level through the first factor (which has a maximum in the intrinsic region), as well as through the second factor (which was discussed in § 3).

(*c*) *Inter-band transitions or trapping: Simple power laws for density of states and transition probabilities; arbitrary disturbance and degeneracy.*

In these cases equations (4.7) and (4.8) may be used to simplify the expressions for the lifetimes. The results, which are easily worked out, will not be given here, as they can become rather complicated in the general case. For example, if it is assumed that the band shapes are those appropriate to a free electron theory,

$$N_e(E_e) = N_{e0}(E_e - E_0)^{1/2}, \quad N_h(E_h) = N_{h0}(E_v - E_h)^{1/2}, \quad \ldots\ldots(5.8)$$

and if it is also supposed that the transition probability is of the form

$$A(E_e, E_h) = A_0(E_e - E_h), \quad \ldots\ldots(5.9)$$

then the lifetime τ_{e_1} of (2.11 *b*) satisfies

$$\frac{1}{\tau_{e_1}} = \frac{\sqrt{(kT)} N_{h0} A_0}{\exp(\delta F_e) - 1}\left\{\Gamma(\tfrac{5}{2}) I(\tfrac{3}{2}, \eta_v - F_0) + \eta_g \Gamma(\tfrac{3}{2}) I(\tfrac{1}{2}, F_0 - \eta_c) - \frac{\tfrac{1}{2}\Gamma(\tfrac{3}{2}) I(\tfrac{1}{2}, F_0 - \eta_c) I(\tfrac{1}{2}, \eta_v - F_0)}{I(-\tfrac{1}{2}, F_0 - \eta_c)}\right\} \quad \ldots\ldots(5.10)$$

5.2. *Interband Transitions or Trapping: Simple Power Laws for Density of States and Transition Probabilities; Small Departure from Equilibrium*

Equations (4.7) and (4.9) may be used to simplify the expressions for the lifetimes. The comments made on case 5.1 (*c*) apply also here.

5.3. *Inter-band Transitions: Energy-Independent Recombination Probability A*

By carrying out an analysis analogous to that for case 5.1 (*a*), it is easily verified that (2.11) again reduces to (2.2) for arbitrary δn and δp. A recombination coefficient exists again and is given by $r = A = \bar{A}$. The assumption of an energy-independent transition probability cannot be applied to the case of electron trapping, because such an assumption would be inconsistent with the argument which leads to equation (3.2).

5.4. *Inter-Band Transitions: Electron Gas Degenerate, Hole Gas Non-degenerate* ($F_0 \gg \eta_c \gg \eta_v$); δn, δp *Otherwise Arbitrary*

At room temperatures, this situation can occur in indium antimonide containing more than 10^{17} electrons per cm^3. It is desirable to define a mean transition probability $\bar{A}'$ of $A(F_0, \eta_h)$, where the averaging process extends over the valence band. It gives the mean probability that an electron from the Fermi surface recombines with a hole:

$$\bar{A}' \equiv \frac{kT}{p_0}\int_{-\infty}^{\eta_v} A(F_0, \eta_h)\, N_h\, P_{h0}\, d\eta_h. \qquad \text{......(5.11)}$$

Lifetimes may be evaluated in this case with the aid of equation (4.10 *a*). The following results are found:

$$\left.\begin{aligned}
a_e &= kT\,N_e(F_0)\left\{\frac{1-\exp(-\delta F_e)}{\delta F_e}\right\}\left\{\frac{2}{1+\exp(-\delta F_e)}\right\},\\
a_h &= \left\{\frac{\exp(-\delta F_h)-1}{\delta F_h}\right\}p_0,\\
B_e &= kTp_0\,\bar{A}'\,N_e(F_0)\left\{\frac{1-\exp(-\delta F_e)}{\delta F_e}\right\}\left\{\frac{2}{1+\exp(-\delta F_e)}\right\},\\
B_h &= n_0 p_0\,\bar{A}\left\{\frac{\exp(-\delta F_h)-1}{\delta F_h}\right\},\\
B_{eh} &= kTp_0\,N_e(F_0)\,\bar{A}'\left\{\frac{1-\exp(-\delta F_e)}{\delta F_e}\right\}\left\{\frac{\exp(-\delta F_h)-1}{\delta F_h}\right\}\\
&\quad\times\left\{\frac{2}{1+\exp(-\delta F_e)}\right\}.
\end{aligned}\right\} \qquad \text{......(5.12)}$$

Using (5.12) in (2.11), it is found that

$$\tau_e = [(p_0+\delta p)\,\bar{A}' + cn_0\,\bar{A}]^{-1}, \qquad \text{......(5.13 a)}$$

$$\tau_h = [(p_0/c+\delta n)\,\bar{A}' + n_0\,\bar{A}]^{-1}. \qquad \text{......(5.13 b)}$$

By (2.6) and (2.8), the net recombination rate is found to be

$$U_{eh} = \left\{\frac{\bar{A}'}{\bar{A}}\frac{\delta n}{n_0}\left[1+\frac{\delta p}{p_0}\right]+\frac{\delta p}{p_0}\right\}R_{eh}. \qquad \text{......(5.14)}$$

For small disturbances, and if $\delta n = \delta p$, this becomes

$$U_{eh} = \left\{\frac{\bar{A}'}{\bar{A}}\frac{1}{n_0}+\frac{1}{p_0}\right\}\delta n\, R_{eh}. \qquad \text{......(5.15)}$$

The analysis for this case shows that the result

$$U_{eh} = (n_0^{-1}+p_0^{-1})\,\delta n\, R_{eh}, \qquad \text{......(5.16)}$$

due to van Roosbroeck and Shockley (1954) does not hold in the present case, unless

$$\bar{A} = \bar{A}'. \qquad \text{......(5.17)}$$

Similarly, it is seen by comparing (5.13) and (2.2) that a recombination coefficient *r* does not exist unless (5.17) is satisfied. This shows that, although (5.16) has been used for the calculation of photon radiative lifetimes in degenerate indium antimonide (Landsberg and Moss 1956), it does not apply exactly in this case. (It should be noted that the average transition rates $\overline{A_{ej}}$, $\overline{A_{hj}}$ are, as an approximation, taken to be independent of the equilibrium Fermi level.)

The quantity $\bar{A}\bar{n}_0/\bar{A}'$ which occurs in (5.15) may be written

$$\int_{-\infty}^{\eta_v} N_h P_{h0}\left[\int_{E_c}^{\infty} A(\eta_e, \eta_h) N_e P_{e0}\, dE_e\right] d\eta_h \Big/ \int_{-\infty}^{\eta_v} N_h P_{h0} A(F_0, \eta_h)\, d\eta_h. \qquad \ldots\ldots(5.18)$$

The integral in square brackets is approximately

$$\boldsymbol{k}T\int_{\eta_c}^{F_0} A(\eta_e, \eta_h) N_e\, d\eta_e. \qquad \ldots\ldots(5.19)$$

One obtains the approximation used by Mackintosh and Allen (1955) for their discussion of the lifetime of added carriers in InSb if (5.19) is replaced by

$$\tfrac{2}{3}\boldsymbol{k}TN_e(F_0)A(F_0, \eta_h) = n_0 A(F_0, \eta_h)/(F_0 - \eta_c). \qquad \ldots\ldots(5.20)$$

Equality of (5.19) and (5.20) implies that the main contribution to the integral comes from an energy range of width $2\boldsymbol{k}T/3$ near the Fermi level and is unlikely to be correct.

5.5. *Trapping: Little Departure from Equilibrium, Electrical Neutrality*

In this case all the terms in equations (3.4) and (3.5), except for the expressions in square brackets, may be evaluated at equilibrium. The terms in square brackets become $\delta F_e - \delta F_t$ and $\delta F_t - \delta F_h$ respectively. Using (3.28) for electrical neutrality, δF_t can be eliminated in favour of δF_e and δF_h. Also, assuming (5.8) to be a good approximation, one finds

$$a_e = n_0 \alpha_e, \qquad a_h = -p_0 \alpha_h, \qquad \ldots\ldots(5.21)$$

where
$$\alpha_e \equiv \frac{I(-\frac{1}{2}, F_0 - \eta_c)}{I(\frac{1}{2}, F_0 - \eta_c)}, \qquad \alpha_h = \frac{I(-\frac{1}{2}, \eta_v - F_0)}{I(\frac{1}{2}, \eta_v - F_0)}.$$

Under these conditions it is found that

$$U_{et} = B_{nn}\delta n + B_{np}\delta_p, \qquad U_{ht} = B_{pn}\delta n + B_{pp}\delta_p. \qquad \ldots\ldots(5.22)$$

The matrix B is given by

$$B = \frac{G}{\sum_{k=0}^{M} Z_k \exp(kF_0)} \times \begin{bmatrix} \left(\dfrac{1}{\alpha_e n_0} + \dfrac{1}{Gt}\right)\sum_1^M (c_{r-1})_0 \exp[(r-1)F_0] & -\dfrac{1}{Gt}\sum_1^M (c_{r-1})_0 \exp[(r-1)F_0] \\ -\dfrac{1}{Gt}\sum_1^M (d_r)_0 \exp(rF_0) & \left(\dfrac{1}{\alpha_h p_0} + \dfrac{1}{Gt}\right)\sum_1^M (d_r)_0 \exp(rF_0) \end{bmatrix} \qquad (\ldots\ldots 5.23)$$

in generalization of equation (A6) of Shockley and Read. The ratio of δn to δp is fixed by requiring (5.22) to apply for a steady non-equilibrium state, so that $U_{et} = U_{ht}$. Hence the lifetimes can be obtained from $\delta n/U$, $\delta p/U$. If one inserts into the final expressions the equilibrium electron and hole generation rates R_{et}, R_{ht} given by (3.21) and (3.22), one finds

$$\left.\begin{aligned} \tau_e &= \frac{\alpha_e n_0}{\alpha_e n_0 + \alpha_h p_0 + tG}\left\{\frac{\alpha_h p_0 + tG}{R_{et}} + \frac{\alpha_h p_0}{R_{ht}}\right\}, \\ \tau_h &= \frac{\alpha_h p_0}{\alpha_e n_0 + \alpha_h p_0 + tG}\left\{\frac{\alpha_e n_0 + tG}{R_{ht}} + \frac{\alpha_e n_0}{R_{et}}\right\}. \end{aligned}\right\} \qquad \ldots\ldots(5.24)$$

If the term tG can be neglected in (5.24), and if the disturbances can be assumed small in (3.23), these two equations go over into each other. If it be assumed that each centre has one trapping state, while an arbitrary degree of degeneracy is allowed, it is instructive to write (5.24) in the form

$$\left.\begin{aligned} \tau_e &= \frac{\alpha_e\alpha_h n_0 p_0/G(d_1)_0 P_{10} + [\alpha_e n_0/G(c_0)_0][\alpha_h p_0/(1-P_{10}) + GP_{10}]}{\alpha_e n_0 + \alpha_h p_0 + GP_{10}(1-P_{10})}, \\ \tau_h &= \frac{\alpha_e\alpha_h n_0 p_0/G(c_0)_0(1-P_{10}) + [\alpha_h p_0/G(d_1)_0][\alpha_e n_0/P_{10} + G(1-P_{10})]}{\alpha_e n_0 + \alpha_h p_0 + GP_{10}(1-P_{10})}. \end{aligned}\right\} \quad \ldots(5.25)$$

Equation (5.25) generalizes the result (A7) of Shockley and Read to an arbitrary degree of degeneracy.

Equations of the above type can be used to discuss the temperature dependence of life times (see, for instance, Ransom and Rose 1954), and the dependence of life times on composition (see, for instance, Burton *et al.* 1953).

REFERENCES

APPEL, J., 1956, *Z. Naturf.*, **11a**, 689.

BURTON, J. A., HULL, G. W., MORIN, E. J., and SEVERIENS, J. C., 1953, *J. Phys. Chem.*, **57**, 853.

LANDSBERG, P. T., 1956, *Proc. Phys. Soc.* B, **69**, 1056 (corrig. 1957, *Ibid.*, **70**, 154).

LANDSBERG, P. T., and MOSS, T. S., 1956, *Proc. Phys. Soc.* B, **69**, 661.

MACKINTOSH, I. M., 1956 a, *Proc. Phys. Soc.* B, **69**, 115 ; 1956 b, *Ibid*, **69**, 403.

MACKINTOSH, I. M., and ALLEN, J. W., 1955, *Proc. Phys. Soc.* B, **68**, 985.

RANSOM, P., and ROSE, F. W. G., 1954, *Proc. Phys. Soc.* B, **67**, 646.

VAN ROOSBROECK, W., and SHOCKLEY, W., 1954, *Phys. Rev.*, **94**, 1558.

SHOCKLEY, W., and READ, W., 1952, *Phys. Rev.*, **87**, 835.

APPENDIX I

OCCUPATION PROBABILITIES OF LOCALIZED LEVELS IN THE PRESENCE OF A MAGNETIC FIELD

If a centre can trap at most one electron the equilibrium occupation probability is (cf. equation (3.1) of the text)

$$p = \frac{1}{1 + Z_0/\lambda Z_1}, \qquad \lambda \equiv \exp F_0. \qquad \ldots\ldots(A1)$$

If the trapping level has total angular momentum quantum number j, it is $(2j+1)$-fold degenerate in the absence of a magnetic field, so that

$$Z_1/Z_0 = (2j+1)\exp(-\eta), \qquad \ldots\ldots(A2)$$

where η is the energy of the trapping level divided by $\boldsymbol{k}T$. In the presence of a field H, this degeneracy is lifted, and

$$\frac{Z_1}{Z_0} = \frac{\sinh(2j+1)z}{\sinh z}\exp(-\eta), \qquad z \equiv \frac{g\beta H}{2\boldsymbol{k}T}, \qquad \ldots\ldots(A3)$$

where g is the splitting factor, and β is the Bohr magneton. For an s-state $j=\frac{1}{2}$, and

$$Z_1/Z_0 = 2\cosh z \exp(-\eta). \qquad \ldots\ldots(A4)$$

If the level can be treated as *non-localized*, however, the appropriate occupation probabilities are the usual ones of the theory of metals.

The results of § 3, together with (A3), yield a theory for the dependence of lifetimes, as limited by trapping, on magnetic field. But no experimental results appear to be available for this effect.

A simple theory of population changes in semiconductors as the result of applying a magnetic field can also be based on (A1) to (A3). The fractional change in electron and hole concentrations thus calculated, contributes a term to the magnetoresistive ratio of semiconductors. Though this term has been considered by Appel (1956), and, under neglect of orbital quantization for the conduction electrons, by Mackintosh (1956 b), these calculations have used

$$1/\{1+\lambda^{-1}\exp(-\eta)\} \qquad \ldots\ldots(A5)$$

for the donor level occupation probability. Use of (A3) instead of (A5) would represent a further refinement of the theory. Its effect on the numerical quantities which occur is most pronounced at low temperatures.

APPENDIX II

TRAPPING: AN ALTERNATIVE MODEL

In § 3 it has been assumed that if a centre is capable of capturing an electron into its ith j-electron state, it is sufficiently accurate to suppose that the centre is in a *particular* $(j-1)$-electron state. This state was, for simplicity, also given the subsidiary label i, and was referred to as the ith $(j-1)$-electron state of the centre. Similarly, it was assumed that if a centre can emit a hole from its ith j-electron state, then the centre can be supposed to be in a *particular* $(j-1)$-electron state, which was referred to as the ith $(j-1)$-electron state of the centre.

It may sometimes be more reasonable to suppose simply that the capture rate of conduction band electrons by j-electron states is proportional to the mean number of centres which contain already $j-1$ electrons. It may then also be assumed that the hole emission rate from j-electron states into the valence band is proportional to the mean number of centres which have captured $j-1$ electrons. This alternative procedure does not affect expressions (3.4) and (3.5) for the capture rates, provided equations (3.6) are replaced by

$$c_{j-1} \equiv Z_{j-1}\int_{E_c}^{\infty} P_e N_e A(E_e, j)\, dE_e,$$

$$d_j \equiv Z_j\int_{-\infty}^{E_v} P_h N_h A(E_h, j)\, dE_h.$$

The rest of the paper is not affected by this change, except that on the right-hand side of equation (3.25) one would have the ratio of the integrals now premultiplied by $(Z_0/Z_1)\exp F_0$.

The Collection of Positive Ions by a Probe Immersed in a Plasma

By J. E. ALLEN†, R. L. F. BOYD‡ and P. REYNOLDS†

† Atomic Energy Research Establishment, Harwell, Berks.
‡ University College London

MS. received 30th August 1956, *and in revised form 28th November* 1956

Abstract. The theory of the collection of positive ions by a probe immersed in a low pressure plasma is reviewed and extended. The ion current is determined mainly by the electric field which penetrates beyond the sheath edge into the plasma. Computed potential distributions for a spherical probe are presented, together with the calculated positive ion current–voltage characteristics.

§ 1. Introduction

The object of the present paper is to review the theory of positive ion collection by a probe and to present computations for the case where (i) the ionic mean free path is large compared with the size of the probe and (ii) the electron temperature is much greater than the positive ion temperature. These are the usual conditions existing in discharge tubes at low pressure.

The classical theory of probe measurements in gas discharges (Langmuir 1923 a, b, c, Langmuir and Mott-Smith 1924, Mott-Smith and Langmuir 1926) contains the basic assumption that the potential difference between a probe and the plasma in which it is immersed, is confined to a space charge region or 'sheath' which surrounds the probe, i.e. it is postulated that the plasma outside the sheath is unperturbed by the presence of the probe. However, during the course of their researches on the low pressure mercury arc, Langmuir and his colleagues found that the potential difference existing between the axis of a discharge tube and the wall was not entirely associated with the wall sheath. It was discovered that the electric field penetrated beyond the sheath edge, into the plasma, and accelerated ions towards the sheath (Tonks and Langmuir 1929). The effect of this field penetration on probe measurements was ignored until the study of Bohm, Burhop and Massey (1949). Recently the subject has been discussed by several writers (Boyd 1950, 1954, Wenzl 1950, Allen and Thonemann 1954).

Bohm *et al.* (1949) showed that the ion current depends on the electron temperature, and not the ion temperature, because the electron temperature determines the strength of the electric field which draws the ions towards the sheath. However, their theory does not deal with the potential distribution within the sheath, nor with the increase in ion current which is observed as a probe is made more negative. The present treatment gives potential distributions both for the plasma and the sheath, together with the positive ion current–voltage characteristics.

The validity of the conventional method of deducing electron temperatures by the retarding field method is unaffected by field penetration.

§ 2. The Orbital Motion Theory

Consider an ion of initial kinetic energy e_2V_0 (where e_2 is the ionic charge†) which describes an orbit round a cylindrical or spherical probe of radius a as shown in figure 1 (inset); in the case of the cylindrical probe e_2V_0 is the energy associated with the velocity component in the plane of the diagram. It is not assumed here that the electric field is restricted to the space charge sheath as in the papers of Mott-Smith and Langmuir (1926). At the distance of closest approach p the kinetic energy of the ion is $e_2(V_0-V)$ where V is the potential ($V<0$) and, as the radial velocity component is zero, the principle of conservation of angular momentum states that

---

† Suffixes 1 and 2 refer to the electron and positive ion respectively.

$$hV_0^{1/2}=p(V_0-V)^{1/2} \qquad \ldots\ldots(1)$$

where h is the impact parameter. Clearly if $p<a$ the ion hits the probe, and ions hitting the probe at grazing incidence have an impact parameter

$$h=a(1-V_P/V_0)^{1/2} \qquad \ldots\ldots(2)$$

where V_P is the potential of the probe. This quantity is the radius of the effective target area of the probe, assuming that all the ions in the plasma have the same initial energy e_2V_0. With this assumption, therefore, the total ion current is given by

$$I_2=AI_{2R}(1-V_P/V_0)^{1/2} \qquad \ldots\ldots(3)$$

for a cylindrical probe, and by

$$I_2=AI_{2R}(1-V_P/V_0) \qquad \ldots\ldots(4)$$

for a spherical probe, where A is the area of the probe and I_{2R} is the positive ion current per unit area in the plasma outside the disturbed region. These calculations can readily be extended to include an initial distribution of ion energies. The expressions obtained for the case of a Maxwellian distribution are†

$$I_2=(4/\pi)^{1/2}AI_{2R}(1-V_P/V_0)^{1/2} \qquad \ldots\ldots(5)$$

and

$$I_2=AI_{2R}(1-V_P/V_0) \qquad \ldots\ldots(6)$$

for the cylindrical and spherical probe, respectively, where V_0 in this case is defined by $e_2V_0=kT_2$.

This theory holds only when the potential distribution satisfies the inequality $V/V_P>(a/r)^2$. This may be shown as follows, referring to figure 1. If the function $h=p(1-V/V_0)^{1/2}$ has a minimum h_M values of h less than this cannot satisfy equation (1). This means that the trajectories of ions of impact parameter $h<h_M$ have no point p at which the radial velocity component is zero. These ions, therefore, continue to move inwards until they hit the probe.

Only the ions of impact parameter $h<h_M$ cross the cylindrical (or spherical) surface of radius r_A (see figure 1) and, as these ions are destined to strike the probe, r_A can be referred to as the 'absorption radius'. Within the absorption radius all the ions move towards the probe and no ions hit the latter at grazing incidence as assumed in the derivation of equations (3) to (6).

It follows from equation (1) that no absorption radius exists if

$$r(1-V/V_0)^{1/2}>a(1-V_P/V_0)^{1/2} \qquad \ldots\ldots(7)$$

or, on rearranging,

$$\frac{V_0-V}{V_0-V_P}>\left(\frac{a}{r}\right)^2 \qquad \ldots\ldots(8)$$

where V_P is the probe potential. If this condition is to hold for ions with a distribution of energies, i.e. if V_0 is very small for some of the ions, then

$$V/V_P>(a/r)^2. \qquad \ldots\ldots(9)$$

---

† Equations (5) and (6) were first derived by Mott-Smith and Langmuir (1926) for the case of a large 'sheath'.

Mott-Smith and Langmuir (1926, p. 750) have given a criterion which can be reduced to (9). In a dense plasma this inequality does not hold and an absorption radius and effective target radius exist, as illustrated in figure 2.

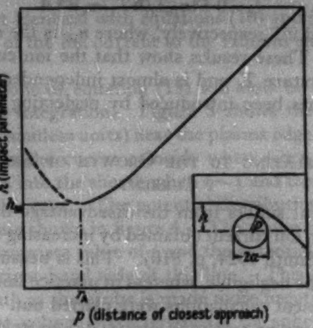


Figure 1. The function $h=p\,(1-V/V_0)^{1/2}$ showing the distance of closest approach of ions which are not captured by the probe.

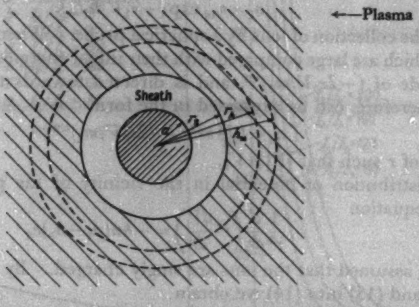


Figure 2. To illustrate the action of a probe. The probe radius, sheath radius, and absorption radius are denoted by a, r_S and r_A respectively; h_M denotes the radius of the effective target area.

§ 3. THE THEORY OF BOHM, BURHOP AND MASSEY (1949)

Bohm *et al.* have calculated the values of h_M for different values of the ion energy ($e_2V_0/kT_1=0{\cdot}01$ and $0{\cdot}5$) for the case of the spherical probe. Clearly the quantity h_M replaces (2) as the effective radius of the probe, because all ions with impact parameters less than h_M are collected by the probe. The 'plasma solution' was obtained using the method of Tonks and Langmuir (1929). This method enables one to calculate the potential distribution in the plasma, where the electron and ion densities (n_1 and n_2) are very nearly equal, but provides no information about the space charge region or 'sheath' which surrounds the probe, so the sheath thickness was assumed to be negligible.

It was found that the potential V_A at the absorption radius r_A was $-2{\cdot}8V_0$ when $e_2V_0/\boldsymbol{k}T_1=0{\cdot}01$ and $-0{\cdot}79V_0$ when $e_2V_0/\boldsymbol{k}T_1=0{\cdot}5$. The calculated ion currents were

$$I_2=0{\cdot}57n_{10}e_2(\boldsymbol{k}T_1/m_2)^{1/2}A \qquad \ldots\ldots(10)$$

$$I_2=0{\cdot}54n_{10}e_2(\boldsymbol{k}T_1/m_2)^{1/2}A \qquad \ldots\ldots(11)$$

for $e_2V_0/\boldsymbol{k}T_1=0{\cdot}01$ and $0{\cdot}5$ respectively, where n_{10} is the electron density in the undisturbed plasma. These results show that the ion current depends mainly on the electron temperature T_1 and is almost independent of the ion energy, so that no serious error has been introduced by neglecting the distribution of ion energies.

§ 4. COMPUTATIONS RELATING TO THE GROWTH OF SHEATH RADIUS AND ION CURRENT

The above treatment suffers from the disadvantage that it does not predict the observed increase in ion current obtained by increasing the (negative) potential (Langmuir and Mott-Smith 1924, p. 810). This is because the thickness of the sheath was assumed to be negligible, whereas in practice this is not always the case. For this reason the present calculations were carried out.

If relatively few electrons reach the probe, due to its negative potential, and the electrons have a Maxwellian energy distribution, then Boltzmann's equation is valid. This equation describes the variation of electron density with potential in the neighbourhood of the probe, viz.

$$n_1=n_{10}\exp(-e_1V/\boldsymbol{k}T_1). \qquad \ldots\ldots(12)$$

Consider the collection of ions by a spherical probe. When the ions have reached energies which are large compared with their initial energies, the ionic velocity has a magnitude of $(-2e_2V/m_2)^{1/2}$, and is directed towards the probe†. The ion density, therefore, can be expressed in the form

$$n_2=I_2[4\pi r^2e_2(-2e_2V/m_2)^{1/2}]^{-1} \qquad \ldots\ldots(13)$$

for values of r such that $|V|\gg V_0$.

The distribution of potential in the vicinity of the probe is described by Poisson's equation

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dV}{dr}\right)=-4\pi(n_2-n_1)e_2 \qquad \ldots\ldots(14)$$

where it is assumed that the ions are singly charged. By substituting n_1 and n_2 from (12) and (13) into (14) we obtain

$$\eta^{1/2}\xi^2\left[\exp(-\eta)+2\left\{\frac{d^2\eta}{d\xi^2}+\frac{2}{\xi}\frac{d\eta}{d\xi}\right\}\right]=\frac{I_2}{I_\lambda} \qquad \ldots\ldots(15)$$

where $\eta=e_1V/\boldsymbol{k}T_1$, $\xi=r/\lambda$, $\lambda=(\boldsymbol{k}T_1/8\pi n_{10}e_1{}^2)^{1/2}$, the Debye distance, and $I_\lambda=(\boldsymbol{k}T_1)^{3/2}/e_2(2m_2)^{1/2}$; the latter is equal to $2(\pi m_1/m_2)^{1/2}$ times the random electron current crossing a Debye sphere (a sphere of radius equal to the Debye distance).

The 'plasma solution' is obtained by neglecting the difference between n_1 and n_2, and with this approximation equation (15) reduces to

$$\eta^{1/2}\xi^2\exp(-\eta)=I_2/I_\lambda. \qquad \ldots\ldots(16)$$

By differentiating equation (16) it may be seen that $d\eta/d\xi\to\infty$ as $\eta\to\frac{1}{2}$, so that the plasma cannot sustain potential differences greater than $\frac{1}{2}\boldsymbol{k}T_1/|e_1|$. Assuming that this potential difference exists between the undisturbed plasma and the sheath edge then the ion current I_2 follows from equation (16) and is given by

---

† It is shown in the appendix that this is true for a spherical probe, but not for a cylindrical probe.

$$I_2 = 4\pi n_{10} e_2 r_s^2 (kT_1/m_2)^{1/2} \exp(-\tfrac{1}{2}) \qquad \ldots\ldots(17)$$

where r_s is the radius of the sheath. Neglecting the sheath thickness this reduces to

$$I_2 = 0{\cdot}61 n_{10} e_2 (kT_1/m_2)^{1/2} A. \qquad \ldots\ldots(18)$$

Equation (18) is almost identical with equations (10) and (11) and this fact illustrates the insensitivity of the ion current to the random energy of the ions in the plasma.

The accurate solutions of equation (15) will now be discussed; these were obtained by numerical integration. Figure 3 shows the variation of potential with distance (in dimensionless units) near the plasma edge. The unit of distance has been chosen so that the curves coincide when $n_1 \simeq n_2$. For large values of I_2/I_λ the plasma merges into the sheath when $\eta \sim \frac{1}{2}$ and the plasma approximation is an excellent one in this case. The potential distributions plotted against ξ, for values of η up to 100, are shown in figure 4. The probe surface corresponds to a vertical line at the appropriate value of ξ and the potential distributions are given by the curves on the right-hand side of this line. This family of curves shows the growth of the sheath as the probe is made more negative. Figure 5 contains the positive ion current–voltage characteristics of the spherical probe, which have been obtained from the data of figure 4, for various values of a/λ.

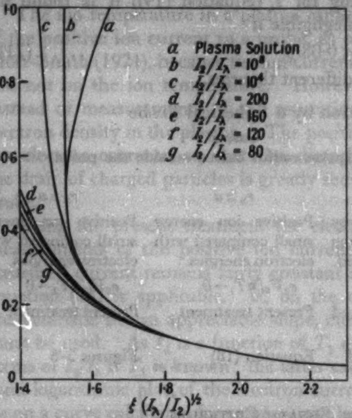


Figure 3. The plasma–sheath transition in the vicinity of a spherical probe.

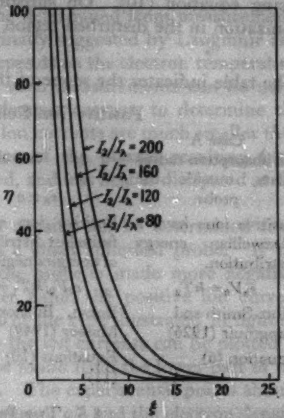


Figure 4. The distribution of potential in the sheath surrounding a spherical probe.

The above treatment neglects ionization in the neighbourhood of the probe and this is valid provided the probe is not too large. Ions created at points where $|V| > V_0$ are collected by the probe, because they have insufficient energy to escape from the potential well, whereas ions created outside the radius at which $|V| = V_0$ have a chance of escaping. For a rough calculation, only the first group of ions will be considered. The ion current flowing to the tube wall, per unit length of discharge tube, is given by the expression $[2\pi n_{10} e_2 r_0 (kT_1/m_2)^{1/2} \exp(-1)]$ where r_0 is the tube radius, because $(e_1 V/kT_1) \sim 1$ and $v_2 \sim (kT_1/m_2)^{1/2}$ at the sheath edge (Tonks and Langmuir 1929). This expression for the ion current gives the ionization rate per unit length of discharge. In the case where $e_2V_0/kT_1=0{\cdot}01$, $|V|=V_0$ at a radius of $3{\cdot}1a$ (see Bohm *et al.* 1949) and the rate of ionization within this radius is given approximately by

$$I_1\sim\frac{4\pi(3{\cdot}1a)^3/3}{\pi r_0^{\,2}}\,[2\pi n_{10}e_2r_0(kT_1/m_2)^{1/2}\exp(-1)]. \qquad \ldots\ldots(19)$$

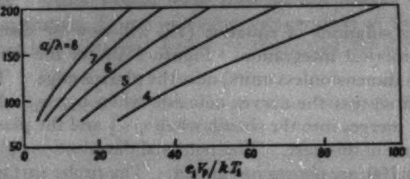


Figure 5. The computed positive ion current–voltage characteristics of a spherical probe.

If the ionization current I_1 is to constitute a negligible fraction of the total ion current to the probe, then

$$I_1\ll 0{\cdot}57n_{10}e_2(kT_1/m_2)^{1/2}4\pi a^2 \qquad \ldots\ldots(20)$$

using equation (10). On substituting for I_1 (equation (19)) it is found that ionization in the disturbed region is negligible if

$$a\ll 0{\cdot}08r_0. \qquad \ldots\ldots(21)$$

The table indicates the scope of the different theories.

Positive Ion Collection by a Spherical Probe

Case A	Case B		
No absorption radius exists, outside the probe	An absorption radius exists, outside the probe		
	$r_s\simeq a$	$r_s\simeq a$	$r_s>a$
Positive ions have a Maxwellian energy distribution.	Positive ion energy finite but distribution of energies neglected.	Positive ion energy small compared with electron energies.	Positive ion energy small compared with electron energies.
$e_2V_0=kT_2$	$e_2V_0/kT_1\neq 0$	$e_2V_0/kT_1\to 0$	$e_2V_0/kT_1\to 0$
Mott-Smith and Langmuir (1926)	Bohm, Burhop and Massey (1949)	Present treatment	Present treatment
Equation (6)	Equations (10) and (11)	Equation (18)	Figures 3–5

§ 5. THE PLANE SHEATH CRITERION

The plasma–sheath transition in the one-dimensional (plane) case has been discussed by Bohm (1949), who neglected the distribution of energies of the ions entering the sheath and assumed that the electric field was zero at the sheath edge. Bohm found that the ion velocity required for sheath formation was

$$v_2=(kT_1/m)^{1/2}. \qquad \ldots\ldots(22)$$

This result has also been obtained by Allen and Thonemann (1954), who avoided the assumption of zero field at the sheath edge. Boyd (1950) has obtained essentially the same result, but his formula is incorrect due to an error introduced in integrating over the velocity distribution of the ions. Also, the plasma cannot be maintained without the creation of new ions (unlike the spherical case treated in

§ 4) and the corresponding terms would have to be included in the equations to obtain a general solution. Tonks and Langmuir (1929) have obtained the approximate 'plasma solution' for a plane discharge and computed the mean ion velocities at the sheath edge. They found that

$$v_2 = 1{\cdot}144(\boldsymbol{k}T_1/m_2)^{1/2} \qquad \ldots\ldots(23)$$

for ion generation proportional to electron density, and

$$v_2 = 1{\cdot}227(\boldsymbol{k}T_1/m_2)^{1/2} \qquad \ldots\ldots(24)$$

for constant ion generation throughout the discharge.

A plane probe, when immersed in a plasma, collects ions from an ill-defined region of roughly hemispherical shape so that, in general, a plane plasma–sheath transition does not exist. However, if we consider the case of a spherical probe in a very dense plasma (i.e. $a/\lambda \gg 1$) the sheath thickness is small and an approximately plane sheath is obtained. If, in addition, the ion temperature is small compared with the electron temperature and if the probe is small so that a plasma is maintained in the vicinity without the creation of new ions (the conditions assumed in § 4) then monoenergetic ions arrive at the plasma edge.

Under these conditions equation (22) is applicable and the resulting expression for the ion current is identical with equation (18).

§ 6. DISCUSSION

The ion temperature in a plasma cannot be determined from measurements of the positive ion current to a probe, as originally suggested by Langmuir and Mott-Smith (1924), because the ion current depends on the electron temperature and not on the ion temperature. However, such measurements can be used, instead of measurements of the saturation electron current, to determine the electron density in the plasma. The positive ion currents are much smaller than the electron currents, so that the disturbance of the plasma resulting from the drain of charged particles is greatly reduced, and less heat is dissipated at the probe.

Using the present treatment the electron density can be determined from measurements of the positive ion current to a small spherical probe. If the measured current remains fairly constant as the probe is made more negative, equation (18) is applicable. If, on the other hand, the positive ion current characteristic has an appreciable slope, curves of the type illustrated in figure 5 must be used. As I_1 is a function of T_1 only the ion currents can be plotted in terms of I_2/I_1 if T_1 is known; the latter can be found from the lower part of the semi-logarithmic plot of the electron current. The experimental points should lie on a curve calculated for some particular value of a/λ and the electron density can then be obtained from the value of λ.

The spherical probe can be suspended by a thin insulated wire so that 'end effects' are practically eliminated. A non-Maxwellian electron energy distribution causes an error in the derived electron density; an example of such a distribution is one in which the drift velocity of the electrons is comparable with their random velocities. Also, secondary electrons emitted from the probe due to the action of metastable atoms, photons, and positive ions contribute to the measured current and cause some error in the electron density (Boyd 1950). A finite ion temperature has very little effect on the ion current, as shown by equations (10), (11) and (18), but ionization in the disturbed plasma surrounding the probe cannot be neglected unless $a \ll 0{\cdot}08 r_0$ (equation (21)). Plasma oscillations have been neglected in this paper, but it is possible that very high frequency oscillations could exist without affecting the motion of the massive positive ions.

ACKNOWLEDGMENTS

The authors wish to thank Dr. P. C. Thonemann for encouraging this work. They are also indebted to Dr. J. Howlett for assistance with the computations.

APPENDIX

ON THE DIRECTION OF THE POSITIVE ION VELOCITY AT RADII OF POTENTIAL $|V| \gg V_0$

For brevity, reference is made to the equations of Bohm, Burhop and Massey (1949), which are denoted thus [ ].

A. Spherical Case

Equation [37] reduces to

$$n_{20}[1-V_A/V_0]r_A{}^2=4n_2r^2(-V/V_0)^{1/2} \text{ if } |V| \gg V_0.$$

The expression for the ion current to the probe is

$$I_2=\frac{n_{20}e_2}{4}\left(\frac{2e_2V_0}{m_2}\right)^{1/2}4\pi r_A{}^2[1-V_A/V_0] \qquad \ldots\ldots[22]$$

and eliminating $n_{20}r_A{}^2[1-V_A/V_0]$ between the above two equations we have $I_2=n_2e_2(-2e_2V/m_2)^{1/2}4\pi r^2$ which shows that the ions are moving radially inwards, because the speed of the ions is $(-2e_2V/m_2)^{1/2}$ if $|V| \gg V_0$.

B. Cylindrical Case

The equation corresponding to [37] is found to be

$$n_2=\frac{n_{20}}{\pi}\sin^{-1}\left[\frac{r_A(1-V_A/V_0)^{1/2}}{r(1-V/V_0)^{1/2}}\right] \qquad \ldots\ldots[37\,a]$$

and the probe current is given by the equation corresponding to [22], viz.

$$I_2=\frac{n_{20}e_2}{\pi}\left(\frac{2e_2V_0}{m_2}\right)^{1/2}2\pi r_A(1-V_A/V_0)^{1/2} \qquad \ldots\ldots[22\,a]$$

but, if $|V| \gg V_0$, $r_A(1-V_A/V_0)^{1/2}=r(-V/V_0)^{1/2}\sin(\pi n_2/n_{20})$ using [37 *a*], so that $I_2=(n_{20}e_2/\pi)(-2e_2V/m_2)^{1/2}2\pi r\sin(\pi n_2/n_{20})$ and therefore $I_2<n_2e_2(-2e_2V/m_2)^{1/2}2\pi r$ since $\sin(\pi n_2/n_{20})<(\pi n_2/n_{20})$. It is seen that the positive ion velocity is not directed towards the probe in this case.

REFERENCES

ALLEN, J. E., and THONEMANN, P. C., 1954, *Proc. Phys. Soc.* B, **67**, 768.

BOHM, D., BURHOP, E. H. S., and MASSEY, H. S. W., 1949, *The Characteristics of Electrical Discharges in Magnetic Fields*, eds Guthrie, A., and Wakerling, R. K. (New York: McGraw-Hill), Ch. 2.

BOHM, D., 1949, *The Characteristics of Electrical Discharges in Magnetic Fields*, eds Guthrie, A., and Wakerling, R. K. (New York: McGraw-Hill), Ch. 3.

BOYD, R. L. F., 1950, *Proc. Roy. Soc.* A, **201**, 329; 1954, *Rocket Exploration of the Upper Atmosphere* (London: Pergamon), p. 336.

LANGMUIR, I., 1923 a, *Science*, **58**, 290; 1923 b, *Gen. Elect. Rev.*, **26**, 731; 1923 c, *J. Franklin Inst.*, **196**, 751.

LANGMUIR, I., and MOTT-SMITH, H., 1924, *Gen. Elect. Rev.*, **27**, 449, 538, 616, 762, 810.

MOTT-SMITH, H., and LANGMUIR, I., 1926, *Phys. Rev.*, **28**, 727.

TONKS, L., and LANGMUIR, I., 1929, *Phys. Rev.*, **33**, 1070, **34**, 876.

WENZL, F., 1950, *Z. angew. Phys.*, **2**, 59.

The Lateral Diffusion of an Electron Swarm in a Magnetic Field

By R. J. BICKERTON
Clarendon Laboratory, Oxford

Communicated by A. von Engel; *MS. received 7th November* 1956

Abstract. The lateral spreading of a swarm of electrons moving through a gas in parallel electric and magnetic fields is measured. Currents of between 10^{-6} and 10^{-10} ampere are passed through helium, neon and hydrogen at pressures of the order 1 mm Hg. Agreement with theory is found for the influence of the magnetic field on the diffusion and mobility of the electrons. The effects of space charge and of secondary processes can also be predicted quantitatively.

---

§ 1. Introduction

The effect of a magnetic field on transport phenomena in ionized gases is of interest in astrophysics, ion-source design, and in discharge physics (Guthrie and Wakerling 1949, Spitzer 1956, Huxley 1937).

When a swarm of electrons moves through a gas in a magnetic field B their diffusion coefficient D_B perpendicular to the field has been shown to be reduced (Townsend 1915) according to

$$\frac{D_B}{D_0} = \frac{1}{1+\omega^2\tau^2} \qquad \ldots\ldots(1)$$

whereas the diffusion coefficient parallel to B remains D_0, viz. that in zero field; $\omega = eB/mc$ is the spiralling frequency of the electrons and τ their mean collision time, equal to l_v/v where l_v is the mean free path for electrons of speed v. Since $\tau \propto 1/p$, where p is the gas pressure, $\omega\tau \propto B/p$, and the similarity laws remain valid in the presence of a magnetic field.

The theory assumes that the current density is such that space charge is negligible, and that $l_v \propto v$, i.e. τ constant and independent of v. For $\omega\tau \gg 1$ the result is largely unaffected by the dependence of l_v on v (Ware 1954, Knoll *et al.* 1934, Tonks and Allis 1937). Experimental confirmation of the theory was obtained (Bailey 1930) using photo-electric currents through hydrogen in a diffusion apparatus of the Townsend type.

The aim of this paper is to investigate the diffusive spreading in combined electric and magnetic fields when ionization in the gas occurs, when secondary processes occur in the diffusion chamber and at higher currents when the effects of space charge are to be included.

§ 2. Theory of the Apparatus

The diffusion chamber (figure 1) has cylindrical symmetry. The electrons from a filament enter through the hole at O, having previously acquired the terminal energy distribution corresponding to the electric field X in a preliminary chamber. They then drift in the z direction in the electric field maintained by guard rings G and at the same time diffuse radially. Finally, they are collected on the disc and ring electrodes A and B. If the current to A is i_A and the total current i_A+i_B then the ratio of these $(R=i_A/(i_A+i_B))$ is a measure of the radial spread of the electron swarm. If space charge and secondary processes in the chamber are neglected then the equation of continuity for the electrons is

$$D\frac{\partial^2 n}{\partial z^2}+\frac{D}{r}\left[\frac{\partial}{\partial r}\left(r\frac{\partial n}{\partial r}\right)\right]-W\frac{\partial n}{\partial z}=0 \qquad \ldots\ldots(2)$$

where n is the number density of electrons at (r, z), D is their diffusion coefficient and W their drift velocity in the z-direction. Solving this equation with the appropriate boundary conditions (Townsend 1915) leads to the following result for the spreading parameter R,

$$R=\frac{i}{i_A+i_B}=\frac{c\left[\dfrac{J_1(\mu_1 a)J_1(\mu_1 c)}{x_1{}^2 J_1{}^2(\mu_1 b)}\exp(-\theta_1 z_0)+\dfrac{J_1(\mu_2 a)J_1(\mu_2 c)}{x_2{}^2 J_1{}^2(\mu_2 b)}\exp(-\theta_2 z_0)+\ldots\right]}{b\left[\dfrac{J_1(\mu_1 a)}{x_1{}^2 J_1(\mu_1 b)}\exp(-\theta_1 z_0)+\dfrac{J_1(\mu_2 a)}{x_2{}^2 J_1(\mu_2 b)}\exp(-\theta_2 z_0)+\ldots\right]} \qquad \ldots\ldots(3)$$

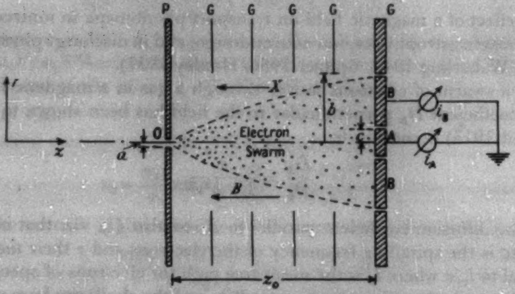


Figure 1. Section of the diffusion chamber.

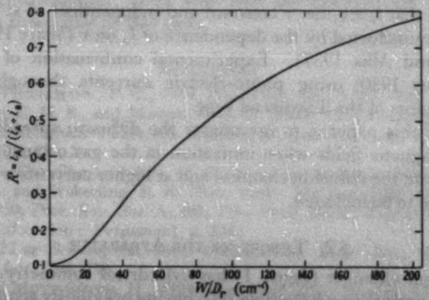


Figure 2. Spreading parameter R as a function of the ratio of drift velocity W to radial diffusion coefficient D_r. $a=0{\cdot}1$ cm, $b=2{\cdot}5$ cm, $c=0{\cdot}5$ cm, $z_0=7{\cdot}9$ cm.

where $\mu_s = x_s/b$, x_s being the sth successive root of $J_0(x) = 0$,

$$\theta_s = (W^2/4D^2 + x_s^2/b^2)^{1/2} - W/2D,$$

z_0 is the length of the diffusion chamber, a is the radius of the hole at O, c is the radius of the disc electrode A, b is the inner radius of the guard rings and also the outer radius of ring electrode B. Thus a relation is obtained between R and W/D for a given apparatus (figure 2), and so by measuring R the ratio W/D can be obtained.

When a magnetic field is applied parallel to the electric field, the drift velocity W and the coefficient D_z for diffusion in the z-direction are unchanged, but the coefficient D_r for radial diffusion is reduced. The coefficient D in (2) described the diffusion in both directions. Since the motion in the z-direction due to diffusion is small compared with that due to the electric field, then D may be associated with radial diffusion only, and can be assumed to vary with the magnetic field in accordance with (1).

The experimental procedure was to measure the variation of the current ratio R with the magnetic field B; X, p and the total current $i_A + i_B$ being kept constant.

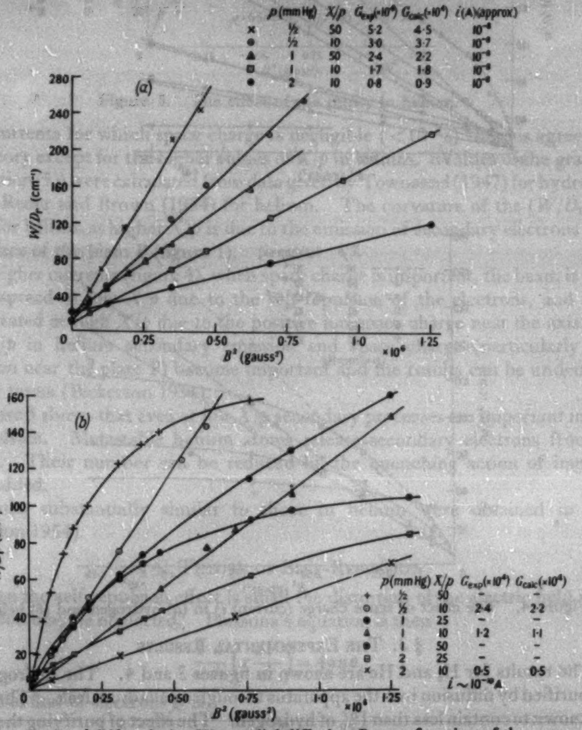


Figure 3. Ratio of drift velocity W to radial diffusion D_r as a function of the square of the magnetic field B for (*a*) hydrogen and (*b*) helium.

From (3) the experimental dependence of W/D_r on B is obtained. Theoretically, equation (1) gives

$$\frac{W}{D_r} = \frac{W}{D_0}\left(1 + \frac{e^2\tau^2}{m^2c^2}B^2\right) \qquad \ldots\ldots(4)$$

and so by plotting W/D_r against B^2 an experimental check of the theory is obtained. The gradient G is, from (4),

$$G = \frac{d(W/D_r)}{d(B^2)} = \frac{W}{D_0}\frac{e^2\tau^2}{m^2c^2}. \qquad \ldots\ldots(5)$$

For constant X/p, $W/D_0 \propto p$, and since $\tau \propto 1/p$, then $G \propto 1/p$.

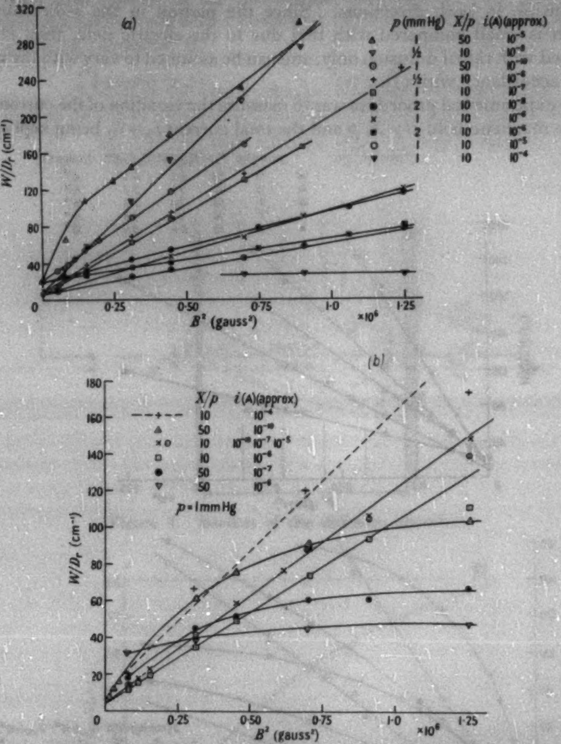


Figure 4. The effect of space charge (current i) in (*a*) hydrogen and (*b*) helium.

§ 3. The Experimental Results

The results for H_2 and He are shown in figures 3 and 4. The hydrogen used was purified by diffusion into the apparatus through a palladium leak. The helium was known to contain less than 1% of hydrogen. The effect of purifying the helium by passing it over activated charcoal cooled to the temperature of liquid oxygen is shown in figure 5.

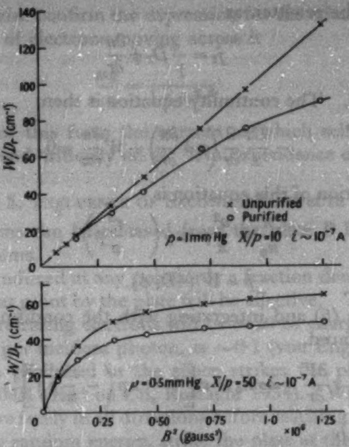


Figure 5. The effect of gas purity in helium.

At currents for which space charge is negligible ($<10^{-7}$A) there is agreement with theory except for the higher values of X/p in helium. Values of the gradient G (equation (5)) were calculated from data given by Townsend (1947) for hydrogen, and by Reder and Brown (1954) for helium. The curvature of the (W/D_r, B^2) results for helium at higher X/p is due to the emission of secondary electrons from the surface of the plate P (figure 1).

At higher currents (figure 4), when space charge is important, the beam is more widely spread at low X/p due to the self-repulsion of the electrons, and more concentrated at high X/p due to the positive ion space charge near the axis. At high X/p in helium secondary emission and space charge (particularly field distortion near the plate P) become important and the results can be understood in these terms (Bickerton 1954).

Figure 5 shows that even at low X/p secondary processes are important in very pure helium. Metastable helium atoms release secondary electrons from the plate P. Their number can be reduced by the quenching action of impurity atoms added.

Results substantially similar to those in helium were obtained in neon (Bickerton 1954).

§ 4. THEORY OF SELF-REPULSION

When the self-repulsion effect is small the distortion of the electric field in the z direction may be neglected. Poissons's equation is then

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dV}{dr}\right)=4\pi ne. \qquad \ldots\ldots(6)$$

The radial flux of electrons j_r is

$$j_r=-D_r\frac{\partial n}{\partial r}+\mu_r n\frac{\partial V}{\partial r} \qquad \ldots\ldots(7)$$

where μ_r is the ' mobility ' of the electrons in the radial direction.

Suppose (7) can be rewritten as

$$j_r = -D_r \phi \frac{\partial n}{\partial r} \qquad \ldots\ldots(8)$$

where ϕ is a constant. The continuity equation is then

$$-\frac{D_r}{r}\frac{\partial}{\partial r}\left(r\phi\frac{\partial n}{\partial r}\right) + W\frac{\partial n}{\partial z} = 0. \qquad \ldots\ldots(9)$$

An approximate solution of this equation is

$$\frac{n}{n_0} = \frac{a^2}{z} p \exp\left(-\frac{r^2}{z}p\right) \qquad \ldots\ldots(10)$$

where $p = W/(4D_r\phi)$.

Substituting for n in (6) and integrating with the condition $\partial V/\partial r = 0$ at $r = 0$, enables $\partial V/\partial r$ to be found.

$$r\frac{\partial V}{\partial r} = 2\pi n_0 a^2 e\left[1 - \exp\left(-\frac{r^2}{z}p\right)\right]. \qquad \ldots\ldots(11)$$

From (7), (8) and (11),

$$\phi = 1 + \frac{\pi a^2 n_0 e \mu_r}{D_r}\left(\frac{z}{r^2 p}\right)\left[1 - \exp\left(-\frac{r^2}{z}p\right)\right] \qquad \ldots\ldots(12)$$

and so if $r^2 p/z < 1$, then

$$\phi = 1 + \frac{\pi a^2 n_0 e \mu_r}{D_r} \qquad \ldots\ldots(13)$$

i.e. ϕ is independent of r for $r^2 < z/p$, so that equation (7) is valid with this condition. If it is assumed that the magnetic field affects the radial mobility and diffusion coefficient equally (Townsend 1915, Huxley 1937) then

$$\frac{\mu_r}{D_r} = \frac{\mu_z}{D_z} = \frac{\mu_0}{D_0} \qquad \ldots\ldots(14)$$

so that with $i = \pi a^2 n_0 e \mu_z X$

$$\phi = 1 + i/D_0 X \qquad \ldots\ldots(15)$$

where i is the total current through the diffusion chamber. Thus self-repulsion becomes important when

$$\frac{i}{D_0 X} = \frac{i}{(D_0 p)(X/p)} \sim 1. \qquad \ldots\ldots(16)$$

In hydrogen at $X/p = 10$ volts cm^{-1} (mm Hg)$^{-1}$ data given by Townsend gives $D_0 p = 10^5$ cm sec^{-1} (mm Hg)$^{-1}$. Here the critical current above which self-repulsion is important $i_c \sim 10^{-6}$ amps. This agrees with our experiments.

For small values of r such that less than half the total number of electrons is enclosed equation (10) is valid, so for a spreading parameter $R < 0{\cdot}5$

$$R = 1 - \exp\left(-\frac{c^2}{z}p\right). \qquad \ldots\ldots(17)$$

This explains the continued linearity of $(W/D_r, B^2)$ when self-repulsion is important. The effect is merely to multiply the values at the abscissa of the $(R, W/D_r)$ curve (2) by the constant ϕ.

The observations also confirm the expression for the effect of a magnetic field B on the mobility μ_B of electrons moving across it

$$\frac{\mu_B}{\mu_0}=\frac{1}{1+\omega^2\tau^2} \qquad \ldots\ldots(18)$$

since if it were not of this form the current at which self-repulsion becomes important would depend critically on B. No dependence on B was observed.

§ 5. INFLUENCE OF SECONDARY EFFECTS

Secondary electrons can be released from the plate P by photons, positive ions or metastable atoms.

Of the photons produced at any point only a fraction determined by the solid angle subtended at that point by the plate will be effective. The photons emitted in helium capable of releasing electrons will be ~20 ev energy, and for these γ_p, the number released per incident photon, is ~0·1 (von Engel 1955).

Each positive ion produced in the space strikes the plate P. The corresponding value $\gamma_i \sim 0{\cdot}01$ (He$^+$ on Cu, Rostagni 1934). Whereas photons and metastable atoms move freely in all directions before reaching P, the ions, because of their relatively low random energy, strike the plate at the same distance from the axis at which they were produced. Since these points of production are determined by the primary swarm, the magnetic field controls to some extent the distribution of secondary electrons.

The metastable atoms of helium have a high γ_m value; figures of the order 0·4 have been measured for the impact of such atoms on copper in vacuum experiments (Oliphant 1929, Dorrestein 1942). Under the conditions of the present work, order of magnitude calculations, using known cross sections (Ebbinghaus 1930, Biondi 1951) show that the number of metastable atoms destroyed in the gas by collisions with electrons or neutral helium atoms is negligible. The metastable atoms are, therefore, all lost by diffusion to the walls.

Calculations for the case $X/p=10$ volts cm^{-1} (mm Hg)$^{-1}$ in helium are now given. Let ϕ be the number of metastable atoms produced by an electron in drifting one centimetre in the direction of the applied field. Using the measured cross sections (Maier-Leibnitz 1935) for excitation to the 2^3S and 2^1S metastable levels (Herzberg 1944) and the theoretical electron velocity distribution (Smit 1936, also Dunlop and Emeléus 1951) it is found that

$$\phi/p=0{\cdot}22.$$

The continuity equation for metastable atoms is

$$-D_m\left[\frac{1}{r}\frac{\partial}{\partial r}\left(\frac{r\partial N}{\partial r}\right)+\frac{\partial^2 N}{\partial r^2}\right]=W\phi n \qquad \ldots\ldots(19)$$

where D_m is the diffusion coefficient for metastable atoms and N is their number density at (r, z). For n, the number density of electrons, the result for the primary beam obtained from equation (2) is used. This assumes that the number density of secondary electrons is everywhere small compared with the density of primaries. Equation (19) was solved for N with the following boundary conditions: $N=0$ at $r=b$ for all z, at $z=0$ for all r, and at $z=z_0$ for all r, i.e. assuming that the inner boundary of the guard rings constitutes the wall, and that this wall and the end plates act as sinks for metastable atoms. From the solution the number of metastable atoms striking unit area of the plate P per second at radius r, viz. $-D_m(\partial N/\partial z)_{r,0}$ is obtained. Each of these incident metastable atoms releases γ_m electrons from P. The radial distribution of released secondary electrons is thus found and this is used as the new boundary condition at $z=0$ for a solution of equation (2). From this solution the ratio R_s between the secondary electron current to the central electrode A and the secondary current to A and B is calculated, as well as the ratio δ between the total secondary and total primary currents.

The calculated values of R_s and δ for various W/D_r are given in the table for helium, with $X/p=10$.

W/D_r	R_s	$\gamma_m\delta/\phi$
40	0·17	1·14
50	0·18	1·16
100	0·20	1·24
150	0·22	1·30

R_s varies slowly with W/D_r, that is with B^2. This results from the ability of the metastable atoms to diffuse freely in the presence of a magnetic field.

By addition the observed current ratio R_{obs} is given by

$$R_{obs}=\frac{R+\delta R_s}{1+\delta} \qquad \ldots\ldots(20)$$

where R is the ratio of currents (equation (3)) when secondary processes are absent. This equation explains the form of the results at high X/p; δ and R_s are almost independent of B, but R rises rapidly as B is increased from zero, thus R_{obs} too rises initially. When R reaches values between 0·7 and 1·0 then W/D_r rises slowly with B and finally approaches a maximum value given by

$$R_{obs}=\frac{1+\delta R_s}{1+\delta}\simeq\frac{1}{1+\delta} \qquad \ldots\ldots(21)$$

independent of B.

From figure 5 $W/D_r=130$ for impure helium and 90 for pure helium at the same value of B^2. If it is assumed that the impurity quenches the metastable helium atoms (Biondi 1951) then the values of R and R_{obs} are 0·65 and 0·52 respectively. With $R_s=0{\cdot}21$ (interpolated from table for $W/D_r=130$) this gives an experimental value $\delta_{exp}=0{\cdot}40$, compared with $\delta=0{\cdot}29$ given by the theory, assuming $\gamma_m=1$ and $\phi/p=0{\cdot}22$. However, the secondary electrons themselves produce metastable atoms which in turn release more secondaries etc., but in the above theory one cycle only has been considered. An approximate extension to include an infinite number of cycles is given by

$$\delta_\infty=(\delta_1+\delta_1^2+\delta_1^3+\ldots)=\delta_1/(1-\delta_1)=0{\cdot}42$$

which is to be compared with $\delta_{exp}=0{\cdot}4$. Although the value of $\gamma_m=1$ is almost certainly too high, in view of the critical dependence of δ_{exp} on the small difference between R and R_{obs} and the approximate nature of the theory the agreement is satisfactory.

The contribution to the secondary current due to the impact on the plate P of positive ions and photons is negligible at $X/p=10$ (Bickerton 1954). At $X/p=50$, positive ions, photons and metastable atoms contribute appreciably to the secondary current.

§ 6. CONCLUSIONS

(*a*) The diffusion of electrons across a magnetic field in a partially ionized gas is shown to be $\propto 1/B^2$.

(*b*) An electron swarm of density such that the self-repulsion is appreciable is used to show that the mobility of electrons across B is $\propto 1/B^2$.

(*c*) Self-repulsion becomes important in a diffusing electron swarm when $i/XD_0 \simeq 1$.

(*d*) Secondary effects in the diffusion chamber must be considered at higher X/p, particularly in rare gases.

(*e*) As a result of (*d*), the experimental method could be adapted to determine the magnitude and relative importance of the secondary processes which precede the electrical breakdown of a gas.

ACKNOWLEDGMENTS

The author wishes to thank Dr. A. von Engel for much assistance, and Professor K. G. Emeléus for criticism. Thanks are due to Lord Cherwell for the facilities of the Clarendon Laboratory.

REFERENCES

BAILEY, V. A., 1930, *Phil. Mag.*, **9**, 560.
BICKERTON, R. J., 1954, *Thesis*, University of Oxford.
BIONDI, M. A., 1951, *Phys. Rev.*, **83**, 1078.
DORRESTEIN, R., 1942, *Physica*, **9**, 433, 447.
DUNLOP, S. H., and EMELÉUS, K. G., 1951, *Brit. J. Appl. Phys.*, **12**, 163.
EBBINGHAUS, E., 1930, *Ann. Phys., Paris*, **7**, 267.
VON ENGEL, A., 1955, *Ionised Gases* (Oxford : Clarendon Press).
GUTHRIE, A., and WAKERLING, R. K., 1949, *Characteristics of Electrical Discharges in Magnetic Fields* (New York : McGraw-Hill).
HERZBERG, G., 1944, *Atomic Spectra and Atomic Structure* (New York: Dover).
HUXLEY, L. G., 1937, *Phil. Mag.*, **23**, 210.
KNOLL, M., OLLENDORF, F., and ROMPE, R., 1934, *Gasentladungstabellen* (Berlin : Springer).
MAIER-LEIBNITZ, H., 1935, *Z. Phys.*, **95**, 499.
OLIPHANT, M. L., 1929, *Proc. Roy. Soc.* A, **124**, 228.
REDER, F. H., and BROWN, S. C., 1954, *Phys. Rev.*, **95**, 885.
ROSTAGNI, A., 1934, *Z. Phys.*, **88**, 55.
SMIT, I. A., 1936, *Physica*, **3**, 543.
SPITZER, L., 1956, *Physics of Fully Ionised Gases* (New York : Interscience).
TONKS, L., and ALLIS, W. P., 1937, *Phys. Rev.*, **52**, 710.
TOWNSEND, J. S., 1915, *Electricity in Gases* (Oxford : Clarendon Press); 1947, *Electrons in Gases* (London : Hutchinson).
WARE, A. A., 1954, *Phil. Mag.*, **45**, 547.

Dipole Moment Fluctuations of a Dielectric Body

By B. K. P. SCAIFE

Electricity Supply Board, Dublin

MS. received 30th August 1956

Abstract. Fröhlich's theory of dielectrics has shown that a dielectric body will exhibit spontaneous fluctuations in dipole moment as a result of thermal agitation. It is pointed out that these fluctuations may arise from a variety of sources and that each source is intimately connected with a particular mechanism of polarization. The particular case of a dielectric which has two mechanisms of polarization (orientational and displacement) is considered. Fröhlich's relations for the total mean square fluctuation in dipole moment of a dielectric sphere are modified so as to yield an expression for the fluctuations arising from only one polarization mechanism (the orientational). It is shown that Fröhlich's free energy method of computing these fluctuations is independent of the dynamic properties of the dielectric. The recent work of Harris and Alder is shown to be in error.

---

§ 1. Introduction

FRÖHLICH (1949) has shown that, in general, a dielectric body will exhibit spontaneous fluctuations in polarization as a result of the thermal agitation of the constituent molecules. He has discussed in detail the case of a dielectric sphere and has shown that, for such a sphere *in vacuo*,

$$(\epsilon_s-1)=\frac{4\pi}{V}\frac{(\epsilon_s+2)}{3}\frac{\langle\mathbf{M}^2\rangle_{\mathrm{av}}{}^{0}}{3kT} \qquad \ldots\ldots(1\,a)$$

and that for a sphere embedded in an infinite volume of its own medium

$$(\epsilon_s-1)=\frac{4\pi}{V}\frac{3\epsilon_s}{(2\epsilon_s+1)}\frac{\langle\mathbf{M}^2\rangle_{\mathrm{av}}{}^{\mathrm{r}0}}{3kT}. \qquad \ldots\ldots(1\,b)$$

ϵ_s is the static dielectric constant, V is the volume of the sphere, k is Boltzmann's constant and T is the absolute temperature. $\langle\mathbf{M}^2\rangle_{\mathrm{av}}{}^{0}$ and $\langle\mathbf{M}^2\rangle_{\mathrm{av}}{}^{\mathrm{r}0}$ are the mean square fluctuations in dipole moment for the two cases just mentioned. The superscript 0 on the angular brackets emphasizes that the external field is zero. The purpose of the superscript r in $\langle\mathbf{M}^2\rangle_{\mathrm{av}}{}^{\mathrm{r}0}$ is to indicate that the reaction field arising from the polarized surroundings of the sphere due to the instantaneous dipole moment $\mathbf{M}$ will affect the value of the average.

By means of a theorem in statistical mechanics (Tolman 1938) Fröhlich has been able to derive (1 *a*) and (1 *b*) using macroscopic quantities only. This theorem states that a system whose Helmholtz free energy, $F(\alpha_1, \alpha_2, \ldots)$, depends upon a number of macroscopic parameters $\alpha_1, \alpha_2, \ldots.$, has the probability

$$\frac{\exp\left[-F(\alpha_1,\alpha_2,\ldots)/kT\right]d\alpha_1\,d\alpha_2\ldots}{\int\exp\left[-F(\alpha_1,\alpha_2,\ldots)/kT\right]d\alpha_1\,d\alpha_2\ldots}$$

of lying in the range $d\alpha_1\,d\alpha_2\ldots.$

In the case under discussion, the macroscopic parameter is $\mathbf{M}$, the instantaneous dipole moment. The details of the derivation of (1 *a*) and (1 *b*) using the theorem just mentioned, have been given by Fröhlich.

It is important to realise that the fluctuations in dipole moment, referred to in (1 *a*) and (1 *b*), are the total fluctuations arising from all causes. In a typical dielectric fluctuations in dipole moment may occur due to a variety of causes, such as the presence of free ions, of molecules with permanent dipole moments, or as a result of intramolecular fluctuations. The question naturally arises as to the possibility of obtaining relations for the fluctuation in dipole moment due to one cause only. For example, there are many dielectrics which have two mechanisms of polarization, namely the orientation of molecules having permanent dipole moments and the displacement of the electrons in a molecule—the so called displacement polarization. It is well known that in electric fields of low frequency both mechanisms are operative in the polarization of such a dielectric It is also known that in fields of sufficiently high frequency the orientational mechanism of polarization becomes inoperative, and only displacement polarization occurs. At low frequencies the behaviour of the dielectric is characterized by the static or low frequency permittivity ϵ_s, and at very high frequencies, beyond the region of dipolar or anomalous dispersion, the behaviour is characterized by a reduced permittivity, ϵ_∞. The quantity ϵ_∞ is usually taken to be approximately equal to the square of the optical refractive index.

A dielectric of the type just described will be considered in the following sections, and relations will be derived for the mean square fluctuation in dipole moment due to the orientational mechanism.

§ 2. Dielectric Sphere in Vacuo

As a result of thermal fluctuation, a dielectric sphere *in vacuo* will have, at some instant, a dipole moment $\mathbf{M}$. This instantaneous dipole moment may be split into two parts $\mathbf{m}_0$ and $\mathbf{m}_d$ such that

$$\mathbf{M}=\mathbf{m}_0+\mathbf{m}_d. \qquad \ldots\ldots(2.1)$$

$\mathbf{m}_0$ is that part of the total dipole moment which is due to the permanent dipole moments of the molecules. Indeed

$$\mathbf{m}_0=\sum_{n=1}^{N}\boldsymbol{\mu}_n \qquad \ldots\ldots(2.2)$$

where $\boldsymbol{\mu}_n$ is the average dipole moment of an isolated molecule and N is the number of molecules in volume V. Thus $\mathbf{m}_0$ is a function only of the orientational coordinates of the molecules. $\mathbf{m}_d$ is the sum of the individual molecular fluctuations in dipole moment and is, consequently, a function of the internal molecular coordinates only.

If we now average over the internal molecular coordinates we find that

$$\langle\mathbf{M}\rangle_{\mathrm{av,\,d}}^{m_0}=\mathbf{m}_0+\langle\mathbf{m}_d\rangle_{\mathrm{av,\,d}}^{m_0}, \qquad \ldots\ldots(2.3)$$

where the suffix d indicates that the average refers to the mechanism of displacement polarization. Note that, because of the polarizing effect of $\mathbf{m}_0$, $\langle\mathbf{m}_d\rangle_{\mathrm{av,\,d}}^{m_0}$ is not zero. The superscript m_0 indicates that the orientational coordinates are held constant.

For a fixed value of $\mathbf{m}_0$, $\mathbf{m}_d$ will fluctuate about its mean value $\langle\mathbf{m}_d\rangle_{\mathrm{av,\,d}}^{m_0}$. If this fluctuation is denoted by $\Delta\mathbf{m}_d$, then

$$\mathbf{M}=\mathbf{m}_0+\Delta\mathbf{m}_d+\langle\mathbf{m}_d\rangle_{\mathrm{av,\,d}}^{m_0} \qquad \ldots\ldots(2.4)$$

because by definition

$$\Delta\mathbf{m}_d = \mathbf{m}_d - \langle \mathbf{m}_d \rangle_{av,\,d}^{m_o}. \qquad \ldots\ldots(2.5)$$

From (2.4) we obtain

$$\mathbf{M}^2 = [\mathbf{m}_o + \langle \mathbf{m}_d \rangle_{av,\,d}^{m_o}]^2 + [\Delta\mathbf{m}_d]^2 + 2\Delta\mathbf{m}_d\,[\mathbf{m}_o + \langle \mathbf{m}_d \rangle_{av,\,d}^{m_o}], \quad \ldots\ldots(2.6)$$

which, in view of (2.3), may be written

$$\mathbf{M}^2 = [\langle \mathbf{M} \rangle_{av,\,d}^{m_o}]^2 + [\Delta\mathbf{m}_d]^2 + 2\Delta\mathbf{m}_d\,\langle \mathbf{M} \rangle_{av,\,d}^{m_o}. \qquad \ldots\ldots(2.7)$$

Since

$$\langle \Delta\mathbf{m}_d \rangle_{av,\,d}^{m_o} = 0 \qquad \ldots\ldots(2.8)$$

we find, on averaging over the internal molecular coordinates, that

$$\langle \mathbf{M}^2 \rangle_{av,\,d}^{m_o} = [\langle \mathbf{M} \rangle_{av,\,d}^{m_o}]^2 + \langle [\Delta\mathbf{m}_d]^2 \rangle_{av,\,d}^{m_o}. \qquad \ldots\ldots(2.9)$$

For clarity $\langle \mathbf{M} \rangle_{av,\,d}^{m_o}$ and $\Delta\mathbf{m}_d$ will be abbreviated to $\mathbf{M}_o$ and to Δ_d respectively, and so

$$\langle \mathbf{M}^2 \rangle_{av,\,d}^{m_o} = \mathbf{M}_o^2 + \langle \Delta_d^2 \rangle_{av,\,d}^{m_o}, \qquad \ldots\ldots(2.10)$$

where $\mathbf{M}_o$ is the average dipole moment of the sphere for a given set of orientational coordinates.

Provided saturation effects are ignored the principle of superposition applies and therefore $\langle \Delta_d^2 \rangle_{av,\,d}^{m_o}$ is independent of m_o; consequently, the superscript m_o will now be discarded.

Averaging next over the orientational coordinates, we conclude that

$$\langle \mathbf{M}^2 \rangle_{av}^{o} = \langle \mathbf{M}_o^2 \rangle_{av}^{o} + \langle \Delta_d^2 \rangle_{av,\,d}, \qquad \ldots\ldots(2.11)$$

where the angular brackets without superscripts refer to averages over all coordinates.

Fröhlich has derived expressions for $\langle \mathbf{M}^2 \rangle_{av}^{o}$ and for $\langle \Delta_d^2 \rangle_{av,\,d}$ so that we can at once obtain an expression for $\langle \mathbf{M}_o^2 \rangle_{av}^{o}$, the mean square fluctuation in dipole moment of the sphere due to changes in the orientational coordinates of the molecules.

Since

$$\langle \Delta_d^2 \rangle_{av\ d} = \frac{3}{2} kT \frac{3V}{2\pi} \frac{(\epsilon_\infty - 1)}{(\epsilon_\infty + 2)} \qquad \ldots\ldots(2.12)$$

and since

$$\langle \mathbf{M}^2 \rangle_{av}^{o} = \frac{3}{2} kT \frac{3V}{2\pi} \frac{(\epsilon_s - 1)}{(\epsilon_s + 2)} \qquad \ldots\ldots(2.13)$$

it follows from (2.11), (2.12) and (2.13) that

$$\langle \mathbf{M}_o^2 \rangle_{av}^{o} = \frac{3}{2} kT \frac{3V}{2\pi} \frac{3(\epsilon_s - \epsilon_\infty)}{(\epsilon_\infty + 2)(\epsilon_s + 2)}. \qquad \ldots\ldots(2.14)$$

§ 3. DIELECTRIC SPHERE EMBEDDED IN ITS OWN MEDIUM

For the case of a dielectric sphere embedded in its own medium, the relation analagous to (2.14) may be derived by an argument similar to that given in § 2.

As in the previous section, the instantaneous dipole moment $\mathbf{M}$ of the sphere will be split into two parts $\mathbf{m}_o$ and $\mathbf{m}_d$. When the average is performed over the internal coordinates we obtain, in place of (2.3), the equation

$$\langle \mathbf{M} \rangle_{av,\,d}^{rm_o} = \mathbf{m}_o + \langle \mathbf{m}_d \rangle_{av,\,d}^{rm_o}. \qquad \ldots\ldots(3.1)$$

$\langle \mathbf{m}_d \rangle_{av,\,d}^{rm_o}$ differs from $\langle \mathbf{m}_d \rangle_{av,\,d}^{m_o}$ because of the reaction field produced by the polarization of the surroundings of the sphere.

For a fixed value of $\mathbf{m}_o$ we may write

$$\mathbf{M} = \mathbf{m}_o + \Delta\mathbf{m}_d^{r} + \langle \mathbf{m}_d \rangle_{av,\,d}^{rm_o} \qquad \ldots\ldots(3.2)$$

where $\Delta^r\mathbf{m}_d$ is defined by the relation

$$\Delta\mathbf{m}_d{}^r = \mathbf{m}_d - \langle\mathbf{m}_d\rangle_{\mathrm{av,\,d}}{}^{rm_o}. \qquad \ldots\ldots(3.3)$$

We deduce from (3.1), (3.2) and (3.3) that

$$\mathbf{M}^2 = [\langle M\rangle_{\mathrm{av,\,d}}{}^{rm_o}]^2 + [\Delta\mathbf{m}_d{}^r]^2 + 2\Delta\mathbf{m}_d{}^r[\langle\mathbf{M}\rangle_{\mathrm{av,\,d}}{}^{rm_o}]. \qquad \ldots\ldots(3.4)$$

By definition

$$\langle\Delta\mathbf{m}_d{}^r\rangle_{\mathrm{av,\,d}}{}^{rm_o} = 0. \qquad \ldots\ldots(3.5)$$

and therefore

$$\langle\mathbf{M}\rangle_{\mathrm{av,\,d}}{}^{rm_o} = \mathbf{M}_{r_o}{}^2 + \langle[\Delta_d{}^r]^2\rangle_{\mathrm{av,\,d}}{}^{rm_o}. \qquad \ldots\ldots(3.6)$$

In (3.6) $\langle\mathbf{M}^2\rangle_{\mathrm{av,\,d}}{}^{rm_o}$ and $\Delta\mathbf{m}_d{}^r$ have been abbreviated to $\mathbf{M}_{r_o}$ and $\Delta_d{}^r$ respectively. $\mathbf{M}_{r_o}$ is the average dipole moment of the sphere for a given set of the orientational coordinates.

We again assume the absence of saturation effects and find that

$$\langle\mathbf{M}^2\rangle_{\mathrm{av}}{}^{r_o} = \langle\mathbf{M}_{r_o}{}^2\rangle_{\mathrm{av}}{}^{r_o} + \langle[\Delta_d{}^r]^2\rangle^r{}_{\mathrm{av,\,d}} \qquad \ldots\ldots(3.7)$$

and since the third member of (3.7) is independent of m_o the superscript m_o has been dropped.

In order to obtain an explicit expression for $\langle\mathbf{M}_{r_0}{}^2\rangle_{\mathrm{av}}{}^{r_o}$, in terms of macroscopic parameters, we must first relate $\langle\Delta_d{}^{r2}\rangle_{\mathrm{av,\,d}}{}^r$ to ϵ_s, ϵ_∞ and T. To do this we must evaluate the Helmholtz free energy of the sphere associated with that part of the displacement polarization $\Delta_d{}^r$.

Following Fröhlich, we divide the free energy into two parts; the first part, F_s, being the self-energy of the sphere *in vacuo* and the other part, F_e, being the interaction energy of the sphere with its surroundings. For F_s we have

$$F_s(\Delta_d{}^r) = \frac{2\pi}{3V}[\Delta_d{}^r]^2\frac{(\epsilon_\infty+2)}{(\epsilon_\infty-1)} \qquad \ldots\ldots(3.8)$$

and for F_e

$$F_e(\Delta_d{}^r) = -\frac{2\pi}{3V}[\Delta_d{}^r]^2\frac{2(\epsilon_s-1)}{(2\epsilon_s+1)}. \qquad \ldots\ldots(3.9)$$

The sum of these two quantities, F_d, is then the Helmholtz free energy associated with the fluctuations in displacement dipole moment $\Delta_d{}^r$. The addition of (3.8) and (3.9) yields

$$F_d = \frac{2\pi}{V}[\Delta_d{}^r]^2\frac{(2\epsilon_s+\epsilon_\infty)}{(\epsilon_\infty-1)(2\epsilon_s+1)}. \qquad \ldots\ldots(3.10)$$

By the theorem quoted in § 1, we find that

$$\langle[\Delta_d{}^r]^2\rangle_{\mathrm{av,\,d}}{}^r, = \frac{\int_0^\infty \Delta^2\exp[-F_d/kT]\Delta^2 d\Delta}{\int_0^\infty \exp[-F_d/kT]\Delta^2 d\Delta}, \qquad \ldots\ldots(3.11)$$

with an obvious abbreviation in notation. It follows immediately from (3.10) and (3.11) that

$$\langle[\Delta_d{}^r]^2\rangle_{\mathrm{av,\,d}}{}^r = \frac{3}{2}kT\frac{V}{2\pi}\frac{(\epsilon_\infty-1)(2\epsilon_s+1)}{(2\epsilon_s+\epsilon_\infty)}. \qquad \ldots\ldots(3.12)$$

In view of the relation (Fröhlich 1949)

$$\langle\mathbf{M}^2\rangle_{\mathrm{av}}{}^{r_o} = \frac{3}{2}kT\frac{V}{2\pi}\frac{(\epsilon_s-1)(2\epsilon_s+1)}{3\epsilon_s} \qquad \ldots\ldots(3.13)$$

substitution of (3.12) and (3.13) into (3.7) produces the equation

$$\langle\mathbf{M}_{r_o}{}^2\rangle_{\mathrm{av}}{}^{r_o} = \frac{3}{2}kT\frac{V}{2\pi}\frac{(\epsilon_s-\epsilon_\infty)(2\epsilon_s+1)^2}{3\epsilon_s(2\epsilon_s+\epsilon_\infty)}. \qquad \ldots\ldots(3.14)$$

We may note here that the inversion of the argument which led to (3.12) would lead to a relation for F_{r_0}, the free energy associated with the orientational fluctuations in dipole moment $\mathbf{M}_{r_0}$. The precise relation is

$$F_{r_0}=\frac{2\pi}{V}\mathbf{M}_{r_0}{}^2\frac{3\epsilon_s(2\epsilon_s+\epsilon_\infty)}{(\epsilon_s-\epsilon_\infty)(2\epsilon_s+1)^2}. \qquad \ldots\ldots(3.15)$$

§ 4. DISCUSSION

The pairs of equations (1 *a*) and (1 *b*), and (2.14) and (3.14) provide us with two similar equations. These equations are:

$$\frac{(\epsilon_s+2)}{3}\langle\mathbf{M}^2\rangle_{\mathrm{av}}{}^{0}=\frac{3\epsilon_s}{(2\epsilon_s+1)}\langle\mathbf{M}^2\rangle_{\mathrm{av}}{}^{r_0} \qquad \ldots\ldots(4.1\,a)$$

$$\frac{(\epsilon\ \ +2)}{3}\frac{(\epsilon_s+2)}{3}\langle\mathbf{M}_0{}^2\rangle_{\mathrm{av}}{}^{0}=\frac{(2\epsilon_s+\epsilon_\infty)}{(2\epsilon_s+1)}\frac{3\epsilon_s}{(2\epsilon_s+1)}\langle\mathbf{M}_{r_0}{}^2\rangle_{\mathrm{av}}{}^{r_0}. \qquad \ldots\ldots(4.1\,b)$$

It is, of course, essential that these two equations are not confused. Equation (4.1 *a*) was deduced by Fröhlich[†]; it relates the total fluctuations in dipole moment of two dielectric spheres, one *in vacuo*, the other embedded in its own medium. The second equation, on the other hand, connects the orientational fluctuations for the two cases. It might be objected that Fröhlich's free energy method for the case of a sphere embedded in its own medium is open to the criticism frequently made of Onsager's theory (Onsager 1936), that in computing the reaction field the static dielectric constant ϵ_s of the surroundings is used and that no allowance is made for the fact that the reaction field may not follow the rapid changes in magnitude and direction of the dipole moment of the sphere. This objection is, in fact, quite groundless. To prove that this is so consider a very large dielectric sphere divided into two regions: one region being a spherical volume situated at the centre of the sphere, the other region being a thick outer shell of dielectric surrounding the inner sphere. The inner sphere is supposed to be sufficiently large for it to have macroscopic dielectric properties and it is further assumed that the volume of the inner sphere is negligible in comparison with the total volume of the sphere. We may write down an expression for the interaction energy between the inner sphere and its surroundings for a given set of coordinates for all the molecules of the whole sphere. We may then average over the coordinates of the molecules in the outer shell only and we shall find that the expression for the interaction energy v of the inner sphere with its surroundings so obtained will be that employed by Fröhlich, namely

$$v=-\frac{4\pi}{3V}\mathbf{M}^2\frac{(\epsilon_s-1)}{(2\epsilon_s+1)} \qquad \ldots\ldots(4.2)$$

in which $\mathbf{M}$ is the instantaneous dipole moment of the sphere. It is to be noted that (4.2) does not depend in any way upon the dynamic properties of the dielectric.

A closer inspection of this aspect of the problem shows that (4.2) is valid provided that saturation effects may be ignored and provided that the sphere may be treated as a point dipole at the centre.

There is one weakness in the derivations of (2.14) and (3.14) and it lies in the omission of consideration of the positional coordinates of the molecules. The fluctuations in these coordinates, so-called translational fluctuations (Kirkwood

---

[†] The writer had been informed that equation (4.1*b*) has been derived in a recent paper by Fröhlich (1956).

1936), will cause fluctuations in dipole moment involving *both* mechanisms of polarization. However it would seem reasonable to include these fluctuations with the orientational fluctuations.

§ 5. CONCLUSIONS

Equation (3.14) has an important application in the macroscopic theory of dielectrics, the reason being that it is unusual to introduce, explicitly, the internal molecular coordinates. Rather, the electronic or displacement polarizability is included by ascribing to each molecule a polarizability α. Such a dielectric will only exhibit fluctuations in dipole moment as a result of fluctuations in the orientational coordinates of the molecules. Reference may be made, however, to the work of Van Vleck (1937) in which the molecules of a dipolar gas are treated as biassed harmonic oscillators.

In view of recent work by Harris and Alder (Harris and Alder 1953, 1954, Harris 1955, Harris and Brush 1956), in which they derive and use a formula for the dielectric constant which is purported to be superior to that proposed by Onsager (1936), it is interesting to note that Onsager's relation follows directly from (3.14) by adopting the appropriate model (see also Fröhlich 1954, 1956).

The model adopted by Onsager and incidentally by Harris and Alder, is well defined and consists in singling out one molecule for consideration. This molecule is assumed to reside in a spherical cavity in a dielectric medium of permittivity ϵ_s. If μ is the dipole moment of the molecule *in vacuo* then (Onsager 1936, Fröhlich 1949)

$$\mathbf{M}_{r0}=\frac{(\epsilon_\infty+2)}{3}\,\frac{(2\epsilon_s+1)}{(2\epsilon_s+\epsilon_\infty)}\mu. \qquad \ldots\ldots(5.1)$$

By combining (3.14) and (5.1) we obtain

$$(\epsilon_s-\epsilon_\infty)=\frac{4\pi N}{V}\,\frac{3\epsilon_s}{(2\epsilon_s+\epsilon_\infty)}\,\frac{(\epsilon_\infty+2)^2}{9}\,\frac{\mu^2}{3kT} \qquad \ldots\ldots(5.2)$$

which is the well known Onsager relation.

The source of error in the work of Harris and Alder which leads them to resurrect the erroneous Piekara–Zakrzewski formula (Piekara and Zakrzewski 1939, Piekara 1950, Davies and Scaife 1954) lies in their use of (4.1 *a*) in place of (4.1 *b*).

ACKNOWLEDGMENT

The writer expresses his best thanks to Dr. Henry J. Crunne for helpful discussions.

REFERENCES

DAVIES, R. O., and SCAIFE, B. K. P., 1954, *J. Chem. Phys.*, **22**, 149.

FRÖHLICH, H., 1949, *Theory of Dielectrics* (Oxford : Clarendon Press); 1954, *J. Chem. Phys.*, **22**, 1804; 1956, *Physica*, **22**, 898.

HARRIS, F. E., 1955, *J. Chem. Phys.*, **23**, 1663.

HARRIS, F. E., and ALDER, B. J., 1953, *J. Chem. Phys.*, **21**, 1031; 1954, *Ibid.*, **22**, 1806.

HARRIS, F. E., and BRUSH, S. G., 1956, *J. Amer. Chem. Soc.*, **78**, 1280.

KIRKWOOD, J. G., 1936, *J. Chem. Phys.*, **4**, 592.

ONSAGER, L., 1936, *J. Amer. Chem. Soc.*, **58**, 1486.

PIEKARA, A., 1950, *Acta Phys. Polon.*, **10**, 37.

PIEKARA, A., and ZAKRZEWSKI, K., 1939, *Nature, Lond.*, **144**, 250.

TOLMAN, R. C., 1938, *The Principles of Statistical Mechanics* (Oxford : Clarendon Press).

VAN VLECK, J. H., 1937, *J. Chem. Phys.*, **5**, 556.

Variations in the Ultra-Violet Absorption of Counting Diamonds

By F. C. CHAMPION and D. L. O. HUMPHREYS
Department of Physics, King's College, London

Communicated by J. T. Randall ; MS. received 9th August 1956, and in revised form 29th November 1956

Abstract. Using a photomultiplier, the ultra-violet absorption of about fifty small octahedral diamonds has been examined. Correlation has been established between counting efficiency and freedom from ultra-violet absorption at various characteristic wavelengths, and the results are tentatively explained in terms of probable defects present in the diamonds.

---

§1. Introduction

In an earlier paper (Champion 1952) a strong correlation was established between the ultra-violet transmission limits of diamonds and the magnitudes of their responses as conduction counters of ionizing radiations. Photographic recording was used for the ultra-violet spectrometer readings and this was often sufficient to demonstrate in addition the existence of narrow absorption bands in a few specimens at some distance from the absorption limit. These preliminary photographic recordings have now been replaced by the use of electrical recording with a photomultiplier, and a few of the earlier results have already been published (Champion 1956). The present paper is a more detailed survey of the ultra-violet absorption spectra of about fifty further diamond counters of similar morphology to those already examined. Meanwhile, Clark, Ditchburn and Dyer (1956) have reported a detailed investigation of the optical absorption of about fifty diamond blocks at 80°K and 290°K, the main instrument being a Hilger Uvispek single beam spectrophotometer. Absolute values for absorption were obtainable since the diamond blocks were ground and polished to give accurately parallel sides for optical transmission. In the present work no absolute measurements were attempted since the diamonds were used in their natural state and the first requirement was for a broad survey of their ultra-violet transmission characteristics for comparison with their efficiency as counters of ionizing radiations. All measurements were made at room temperature.

§ 2. Experimental

A Hilger quartz spectrograph, type E498, was used in conjunction with two light sources, one being a hydrogen lamp with a quartz envelope, for the range 2200Å to 3410Å and the other a tungsten lamp for the range 2914Å to 4930Å.

The diamonds were placed between two brass plates forming a holder which could be slid on to a plate fixed to the front of the spectrograph, the central portion of the crystals coinciding with two small holes in the plates: the diameter of the holes was small relative to the dimensions of the diamonds. The absorption spectrum of a diamond was examined using a photomultiplier which could be moved slowly by means of a motor across the focal plane of the spectrograph.

All the traces were taken under standard conditions, one of the wavelength ranges being 2474Å to 3410Å. A suitable width of the spectrograph slit was found using some of the diamonds and then adopted for all diamonds. Diamonds transparent below 2500Å were again examined similarly adopting a standard wavelength range, 2200Å to 2692Å and a standard width of slit. The adoption of standard ranges, and to each of these a standard spectrograph slit width, provided the means of examining relatively the absorption spectra of diamonds. No attempt was made to determine the absolute absorption of the diamonds at different wavelengths as the characteristic features of the absorption spectra were already clearly indicated in the traces obtained.

The diamonds examined were mostly natural octahedra, and all except one possessed parallel faces. Care was taken in setting the diamonds in the holder so that the faces were parallel to the plates and also in placing the crystals centrally with respect to the two holes of the holder plates.

With the rigorous precautions outlined above it was found that many reproducible results could be obtained in a day. No effect on the traces was found if a diamond was removed and replaced between readings, and this reproducibility was found even after five months; the peaks which characterized some of the spectra were much more prominent than any variations in the characteristics of the equipment, the photomultiplier response or in the spectral properties of the hydrogen discharge tubes. Errors arising from any mis-settings of the diamond will only affect the overall intensity of the trace, the possible variations being almost independent of the wavelength of the light, and therefore no spurious absorption peaks would be expected.

The crystal which possessed non-parallel faces was a cleaved portion of a diamond, the other portion possessing parallel faces. The traces of both portions were very similar and this could be interpreted as indicating that, having taken reasonable care, the traces obtained for all the diamonds examined are valid for comparative examination.

The accuracy to which the wavelength of a peak could be estimated on a particular trace ranges from ± 6Å at the wavelength 2650Å to ± 12Å for wavelengths greater than 3200Å. These estimates take into account the accuracy of the calibration charts as well as the accuracy by which the absorption spectrum traces could be read.

§ 3. Results

Over fifty traces were obtained (Humphreys 1955), figure 1 (*a*) showing the intensity against wavelength recording of the hydrogen lamp source in the range 2000 Å–2700 Å with diamonds absent, and figures 1 (*b*), 1 (*c*) and 1 (*d*) showing representative traces when certain diamonds are present. The cut-off is defined as the wavelength at which the response of the photomultiplier is indistinguishable from the noise level. Trace 1 (*b*) shows a cut-off at 2264 Å and a prominent absorption band at 2360 Å, while trace 1 (*c*) also shows a cut-off at 2264 Å but no absorption band at 2360 Å. It is obviously impossible to determine the presence of absorption bands at wavelengths shorter than the cut-off wavelength and consequently for specimens with traces such as 1 (*d*), such investigations were confined to wavelengths longer than its cut-off at 3004 Å. For this specimen a prominent absorption occurs at 3157 Å, where it is also just detectable on the complete trace of 1 (*c*) but absent from 1 (*b*). In some other specimens the bands already referred to appear as points of inflection in the trace, and further bands of much feebler intensity than the above may be present at other wavelengths. The cut-off wavelengths of the present specimens cover the range 2250 Å through intermediate values to 3063 Å.

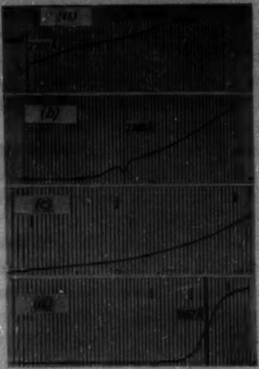


Figure 1.

The counting response covers the range 8400 ion-pairs-cm (see Taylor 1956) down to the noise level of the recording apparatus at 300 ion-pairs-cm, the specimens responsible for traces 1 (*b*), 1 (*c*) and 1 (*d*) giving counting responses of 1200, 8400 and 550 respectively. The birefringence was fairly strong for the first specimen and weak for the other two. These properties are summarized in the table.

Spectral trace type	λ cut off	Absorption bands at		Counting	Birefringence
		$\lambda = 2360$	$\lambda = 3157$		
1 (*b*)	2264	Prominent	None	1200	Strong
1 (*c*)	2264	Very weak	Weak	8400	Weak
1 (*d*)	3004	—	Prominent	550	Weak

Similar analyses were made for all the specimens and a classification may be attempted according to various spectral and counting properties. It is first essential to clarify the situation regarding the designations of type 1 and type 2 specimens, where some confusion has arisen in the literature. Some workers refer to specimens as type 1 if their transmission is undetectable by present apparatus below 3000 Å, whereas others refer to all specimens as type 1 unless they show no detectable absorption bands or edges at wavelengths greater than about 2250 Å when they are designated as type 2. Such a criterion if used by different workers is clearly quite arbitrary, for the ability to detect the presence or absence of an absorption band is a function of the sensitivity and resolution of the detector, just as is the detection of the counting properties or any other physical property. Thus, Clark, Ditchburn and Dyer (1956) defined type 1 diamonds as those showing a secondary absorption edge at 3320 Å, whereas type 2 diamonds showed only the fundamental absorption edge at 2250 Å. Nevertheless, mixed and intermediate types were found which in some cases showed little evidence of the secondary absorption edge but which showed the 2360 Å band very prominently. In the present experiments the transmission trace of one specimen shows no appreciable irregularities except a most pronounced absorption edge at 3064 Å which might logically be referred to as a tertiary absorption edge. To avoid such complications the terms type 1 and type 2 will not be used; instead, the present results will be considered in terms of spectroscopic groups, group 1 comprising those specimens whose cut-off as measured with the present apparatus was above 3000 Å, and group 2 those whose cut-off determined by the same apparatus was below 3000 Å. As shown in figure 2, where the magnitude of the counting response is given by the length of the vertical line, the specimens being arranged in order of increasing wavelength of cut-off, on this classification twelve were group 1 specimens and the remainder group 2. The mean counting pulse height for the group 2 specimens was about 3000 ion-pairs-cm, while for the group 1 specimens it was barely detectable at the noise level of 300 ion-pairs-cm.

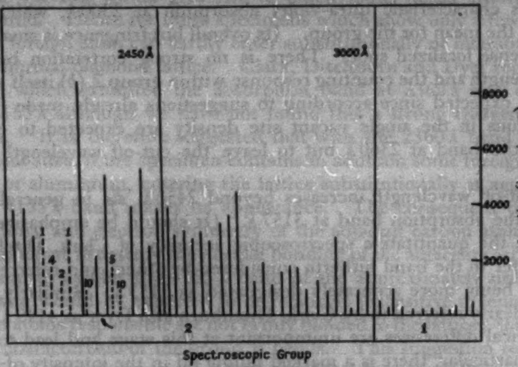


Figure 2.

§ 4. Discussion

The counting properties of these specimens agree with previous results on other diamonds in that they show that on the average, group 2 diamonds are better counters than group 1 specimens by at least an order of magnitude. A more detailed examination of figure 2, however, exhibits the wide range of variety of behaviour of which diamonds are capable. For all seventeen specimens whose ultra-violet cut-off was shorter than 2450 Å, the absorption band at 3157 Å was absent or very weak. Seven of these gave traces like figure 1 (*b*) while in the remaining ten the absorption band at 2360 Å was absent as in figure 1 (*c*) or existed only as a point of inflection. The mean counting response of the first set to be denoted as group 2 (*b*), and shown as dotted lines in figure 2, was 1700 ion-pairs-cm and of the latter was over 4000 ion-pairs-cm. This correlation implies that the defect responsible for the absorption band at 2360 Å is also responsible for reducing the counting efficiency and it has been suggested that this defect is associated with single vacant sites (Champion 1956). While, therefore, on the spectral evidence alone the specimens could be placed in groups 2 (*b*) and 2 (*a*) according as the absorption band at 2360 Å was very prominent or barely detectable, on defect theory they form a single group differing only in the single vacant site density. The best counters of all the present specimens gave a response of 8400 ion-pairs-cm, which is more than twice the average of those in group 2. Its ultra-violet trace is shown in figure 1 (*c*) and shows a cut-off at 2264 Å and no detectable absorption bands. It is also free from birefringence and thus satisfies all tests so far made for a high degree of freedom from defects of all types.

The numbers beside the dotted lines in figure 2 give the approximate relative intensities of the absorption at 2360 Å. The worst counter of group 2 (*b*) has a response which is about half that of the mean of its group. This value correlates well with its characteristic ultra-violet absorption at 2360 Å, which is about twice that of the mean for the group. Its overall birefringence is small although it has an intense localized spot. There is no strong correlation between the cut-off wavelength and the counting response within group 2 (*b*) itself nor is such a correlation expected since according to suggestions already made (Champion 1956) variations in the single vacant site density are expected to change the intensity of the band at 2360 Å but to leave the cut-off wavelength relatively unaffected.

As the cut-off wavelength increases beyond 2450 Å, so, in general, does the intensity of the absorption band at 3157 Å. (It should be emphasized that we cannot claim the quantitative spectroscopic accuracy of Clark, Ditchburn and Dyer who refer to the band hitherto considered by others, and referred to by us as 3157 Å, as being more accurately designated as 3153 Å. Similarly, the band at 4155 Å is clearly referred to by Clark, Ditchburn and Dyer as 4149 Å, but these slight numerical differences are unimportant at this stage and lead to no ambiguity.) In particular, there is a marked falling off in the intensity of the 3157 Å band as the cut-off approaches about 3000 Å and above. At the same time, a further absorption band sometimes becomes prominent at 4155 Å together with the appearance of other fainter absorption bands. It is appropriate at this point to draw a comparison with the results of other workers since all the bands mentioned have already been reported for quite different specimens (Sutherland, Blackwell and Simeral 1954, Clark *et al.* 1956), together with many other bands not investigated here. These workers have shown that the absorption bands may be divided into two groups, group A being exemplified by the 3157 Å band and group B by the 4155 Å band. In particular, the intensity of group A bands had been shown to be related to the cut-off wavelength and to be temperature sensitive, whereas the intensity of the group B bands was independent of both cut-off wavelength and of temperature. It may, therefore, be inferred that these two groups of bands arise from different defects and it has been suggested (Champion 1956) that the 3157 Å band and the cut-off are due to aggregate vacancies or cavity sites. On the other hand, the 4155 Å band is well-known to correspond to the blue fluorescence produced by ultra-violet excitation in some diamonds and by ionizing radiations in all diamonds. It has been long suspected (Bull and Garlick 1950) that the 4155 Å line is in some way related to the transition $^1S_0$–$^3P_1$ or $^1S_0$–$^3P_2$ present in the *isolated* carbon atom but modified by the neighbourhood of the surrounding atoms in the solid. Such transitions are, of course, forbidden in perfect diamond since they would imply levels in the forbidden energy zone. It seems, therefore, not unreasonable to suppose that its appearance implies that for some of the carbon atoms present the tetrahedral bonds must be broken either momentarily or more or less permanently. The passage of an ionizing particle will obviously achieve this momentarily and thus account for the blue fluorescence of all diamonds under ionizing radiations. For its presence in ultra-violet absorption as found in certain diamonds, vacant sites or aggregate vacant sites are barely sufficient disturbance since the carbon atoms round the site may still be supported by three of the four normal carbon bonds. Even the fourth, or defect bond, although weaker than the normal bond, is shared symmetrically around the vacant site. A more serious defect is therefore required such as would be provided by the presence of a foreign atom. This disturbance, however, must not be so serious as to produce large numbers of levels in the forbidden gap for this would give rise to much more varied absorption than is actually found. Hence for those specimens which show only a few well-defined bands, the foreign atom can hardly enter substitutionally or interstitially into the otherwise normal diamond lattice. Clark, Ditchburn and Dyer report that their specimens which exhibited the absorption band at 2360 Å also exhibited the band at 4155 Å although we have not found this a strong correlation with our specimens. It is therefore suggested that with the 2360 Å band attributed to single vacant sites, if the specimen contains in addition some foreign atoms, such as silicon or aluminium, entering the lattice substitutionally at aggregate vacant sites, the 4155 Å band makes its appearance. The consequent disturbance of the vibrational and electron symmetry of the adjacent carbon atoms might then become sufficient to destroy the normal bonding of the adjacent carbon atoms and permit them to exhibit the 4155 Å band of the approximately unbonded atom. Such a band would be expected to be relatively insensitive to temperature since the carbon atoms responsible are not firmly bonded to the lattice; this insensitivity is indeed characteristic of the group B bands. This suggestion receives further support from the observations of Clark, Ditchburn and Dyer on the bands produced by irradiation of diamonds where interstitial carbon atoms and vacant sites are produced. These bands, while not in exactly the same position as the group B bands of natural non-irradiated diamonds, show marked similarity in their appearance and in their insensitivity to temperature, thus indicating that such bonds are indeed due to carbon atoms which are not properly bonded to the main lattice. It seems clear that spectroscopic data will provide the main basis of diamond analysis, the choice of interpretations being guided by the need for consistency with interpretations from other techniques.

ACKNOWLEDGMENTS

We are grateful to Professor W. C. Price, King's College, London, for advice on the spectrometer and to Mr. K. Taylor for assistance with the determination of the counting efficiencies of the diamonds. One of us (D. L. O .H.) is indebted to the Montgomeryshire Education Committee for financial assistance.

REFERENCES

BULL, C., and GARLICK, G. F. J., 1950, *Proc. Phys. Soc.* A, **63**, 1283.
CHAMPION, F. C., 1952, *Proc. Phys. Soc.* B, **65**, 465; 1956, *Proc. Roy. Soc.* A, **234**, 541.
CLARK, C. D., DITCHBURN, R. W., and DYER, H. B., 1956, *Proc. Roy. Soc.* A, **234**, 363.
HUMPHREYS, D. L. O., 1955, *M.Sc. Thesis*, University of London.
SUTHERLAND, G. B. B. M., BLACKWELL, D. E., and SIMERAL, W. G., 1954, *Nature, Lond.*, **174**, 901.
TAYLOR, K., 1956, *Ph.D. Thesis*, University of London.

RESEARCH NOTES

High Temperature Susceptibilities

By A. DANIELIAN and K. W. H. STEVENS
Department of Physics, University of Nottingham

Communicated by L. F. Bates ; MS. received 12th November 1956

As a result of recent susceptibility measurements on manganous sulphide (Carter and Stevens 1956) we have been led to study the high temperature properties of antiferromagnetics in which the magnetic ions are on a face-centred cubic lattice with nearest neighbour interactions only. The main problem is to extract the value of the exchange interaction from a knowledge of the Néel temperature (if it exists) and the asymptotic Curie temperature.

With regard to the existence of a Néel temperature, Li (1951) and Ziman (1953) conclude that the system will not show antiferromagnetism. We feel that this conclusion may be misleading, for what they really show is that there will be no superlattice structure. It seems that there may still be anomalies in the susceptibility and specific heat at some finite temperature, though these may not be as sharp as in a system in which a superlattice is formed. Whether this is so or not it would seem that the use of any approximation involving long range order will lead to difficulties, and we have therefore been examining the method of high temperature power series (Van Vleck 1937).

The essence of this is that the susceptibility χ is expressed as a power series in J/kT,

$$\chi \equiv C\left(\frac{J}{kT}\right)\left[1+a_1\left(\frac{J}{kT}\right)+a_2\left(\frac{J}{kT}\right)^2+\dots\right]$$

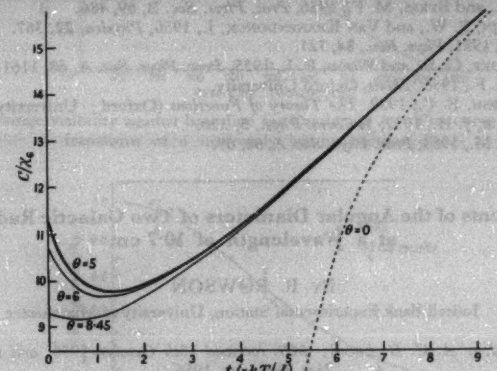
where the interaction between a pair of neighbouring spins, $\mathbf{S}_1$ and $\mathbf{S}_2$, is $-2J\,\mathbf{S}_1 \,.\, \mathbf{S}_2$. In principle all the a's can be calculated, but in practice this is very difficult, and for $S=\frac{1}{2}$, the case on which we shall concentrate, the a's are known only up to a_6 (Domb and Sykes 1956). Consider now the power series:

$$f(z)=1+a_1z+a_2z^2+\dots.$$

where z is a complex variable. Starting from the origin and moving along the positive real axis its value determines the susceptibility at T, where $kT=J/z$. This implies that J is positive, or that a ferromagnetic system is being considered. Similarly the negative real axis gives the susceptibility of an antiferromagnetic as a function of temperature, so that, by considering the properties of $f(z)$, ferro- and antiferromagnetism can simultaneously be studied. Quoting now some results about power series (Titchmarsh 1939) it is well known that every power series has a circle of convergence, inside which the infinite sum is convergent and outside which it is not. The radius of this circle is determined by the position of the first singularity. Also, any power series defines an analytic function within a natural boundary, so that by analytical continuation the function can, in general, be continued outside the circle of convergence of the power series. These ideas are used physically by seeking the singularities of the analytic function $f(z)$ which lie on the real axis. The one nearest to the origin on the positive side, F, corresponds to the ferromagnetic Curie temperature, and the nearest one on the negative side, A, to the Néel temperature.

It is usually assumed that all the *a*'s are positive (all the known ones are) and under these circumstances the singularity nearest to the origin lies on the positive real axis and is therefore F. A good deal of effort has been devoted to estimating its position (Rushbrooke and Wood 1955, Sykes 1956). If now one moves out from the origin along the negative real axis the series will certainly diverge outside the circle determined by F, and will no longer represent the antiferromagnetic susceptibility even though the temperature may be well above the Néel temperature. Clearly a means of analytically continuing the power series in the direction of the negative real axis is required. There are many formal ways of doing this, but many are ruled out on practical grounds, because if $w=f(z)$, and a power series in w is used, it is essential that the initial coefficients in the w series shall depend only on the initial coefficients in the z series. For example, a simple displacement $w=z+\alpha$ cannot be used. An obvious transformation to use is the bilinear one, $1/w=\theta+1/z$, for it satisfies the above condition and, by choosing θ suitably, F can be projected towards infinity while A remains at a finite point. Also, the expression for the susceptibility takes on a form very similar to the Curie–Weiss law, being a power series in $\theta+1/t$, where $t=-kT/J$.

In the figure we have plotted C/χ_6 against t, where χ_6 denotes the susceptibility expression truncated at six terms, for various values of θ. The curve for $\theta=0$ is that which would be obtained without using the transformation, and it begins to go very wrong for t less than about 7, showing the influence of the radius of convergence, for which the corresponding value of t is approximately 4·2. The other curves behave much more like experimental susceptibility curves down to at least $t=3$. If all the *a*'s were known, instead of just six, these curves would coincide, so that their divergence gives some measure of the error made in the truncation. It therefore seems likely that down to $t=3$ we have an accurate representation of the true susceptibility. An important feature of the curves is that they show a fairly steady curvature, which means that considerable care is necessary in deducing the values of C, the Curie constant, and of the asymptotic Curie



temperature. It seems likely that the minima in the curves occur in the neighbourhood of the Néel temperature, and we may tentatively estimate that this will not be above $t=1{\cdot}2$. This means that the ratio of the Néel temperature to the asymptotic Curie temperature is less than 0·2, which is much less than the corresponding ratio for the simple cubic and body-centred cubic lattices (Kasteleijn and Van Kranendonck 1956). (It may be noted that Kasteleijn and Van Kranendonck and Li on the one hand and Rushbrooke and Wood on the other cannot all be correct, for the A's of Kasteleijn and Van Kranendonck and Li, for the simple cubic and body-centred cubic lattices, fall inside the corresponding circles of convergence of Rushbrooke and Wood.) It seems that to obtain the correct asymptotic Curie temperature by extrapolation of the $(1/\chi, T)$ curve it is necessary to use experimental data obtained at temperatures at least ten times the Néel temperature, and we believe that the failure to do this may account for both the high value of the asymptotic Curie temperature obtained in MnS, and for some error in the magneton number.

We are applying this method to other lattices. Preliminary results suggest that in the s.c. and body-centred cubic lattices the points F and A are at almost the same distance from the origin, so that the extension of $(1/\chi, T)$ is not so striking.

References

Carter, W. S., and Stevens, K. W. H., 1956, *Proc. Phys. Soc.* B, **69**, 1006.
Domb, C., and Sykes, M. F., 1956, *Proc. Phys. Soc.* B, **69**, 486.
Kasteleijn, P. W., and Van Kranendonck, J., 1956, *Physica*, **22**, 367.
Li, Y. Y., 1951, *Phys. Rev.*, **84**, 721.
Rushbrooke, G. S., and Wood, P. J., 1955, *Proc. Phys. Soc.* A, **68**, 1161.
Sykes, M. F., 1956, *Thesis*, Oxford University.
Titchmarsh, E. C., 1939, *The Theory of Functions* (Oxford : University Press).
Van Vleck, J. H., 1937, *J. Chem. Phys.*, **5**, 320.
Ziman, J. M., 1953, *Proc. Phys. Soc.* A, **66**, 89.

Measurements of the Angular Diameters of Two Galactic Radio Sources at a Wavelength of 10·7 cm

By B. ROWSON
Jodrell Bank Experimental Station, University of Manchester

Communicated by A. C. B. Lovell; MS. received 30*th October* 1956, *and in final form* 21*st December* 1956

The apparent angular diameters of the most intense radio sources were successfully measured at metre wavelengths in 1952 (Smith 1952, Mills 1952, 1953, Hanbury Brown *et al.* 1952, 1955). Subsequently, these sources were found to give measurable intensities at centimetre wavelengths (Haddock *et al.* 1954) but as yet no measurements of angular diameters have been made in this range. The present note describes the preliminary results of measurements of the angular diameters of the Cassiopeia (right ascension 23 h 21 min 12 sec Dec. +58° 32·1′) and Taurus (right ascension 05 h 31 min 30 sec Dec. +22° 3′) radio sources on a wavelength of 10·7 cm.

The apparatus consisted of an interferometer using two focal plane paraboloidal aerials of 10 ft and 25 ft diameter respectively. Each aerial was fitted with its own primary feed, waveguide mixer, klystron local oscillator and I.F. preamplifier. The outputs of the two intermediate frequency preamplifiers were taken by equal lengths of coaxial cable to a main intermediate frequency amplifier where the two signals were added together, further amplified and detected with a square law detector. The output of the detector was divided into two parts. The d.c. component was taken to a total power recorder, while the video component was taken to a phase sensitive detector, the output of which was displayed on a second recorder which recorded the interference fringes.

C.W. signals at about 2800 Mc/s were taken from each klystron through coaxial cable to a coaxial crystal mixer. The output of this mixer was used to tune one klystron to keep the difference frequency between the two local oscillators equal to 22 ± 1 kc/s and also to provide the reference voltage for the coherent detector. In principle this interferometer was similar to that described by Hanbury Brown, Palmer and Thompson (1955).

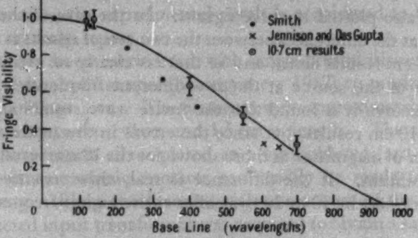


Figure 1. Fringe visibility against base line for Cassiopeia A. The curve is the Fourier transform of a uniformly bright circular disc.

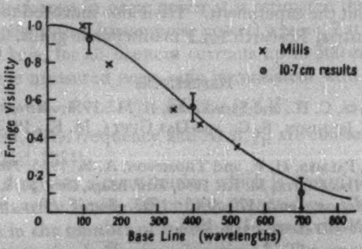


Figure 2. Fringe visibility against base line for Taurus A. The curve is the Four transform of an integrated strip source with gaussian E-W intensity distribution.

As far as the total power record was concerned the apparatus consisted of two straight receivers with their outputs added together. With this arrangement it was not possible to measure the total power received from a source because of variations in receiver noise level. Hence it was not possible to measure the fringe visibility absolutely and so the fringe amplitudes on the fringe record, corrected for variations in receiver gain, were taken to be proportional to the fringe visibility. This was based on the assumption that the aerial gains did not change throughout the experiment. To check this, the first measurements made on Cassiopeia A were repeated at the end of the experiment.

Measurements were made on a number of E–W base lines of different lengths up to 700 wavelengths. The results normalized to give unit fringe visibility at zero base line, are shown in the table and are plotted in figures 1 and 2. For comparison the corresponding metre wave results derived from papers by Smith (1952) and Jennison and Das Gupta (1956) for Cassiopeia, and Mills (1953) for Taurus are also plotted in these figures. In the case of the Taurus source it can be seen that the differences between the two sets of results is smaller than the errors in the 10 cm results alone, and so there is clearly no significant difference in the diameter of the source at the two different frequencies. However, for the Cassiopeia source it is found that the metre wave results are systematically lower than the 10 cm results, but since the errors in the metre wave results are of the same order of magnitude as those shown for the 10 cm results, the difference may not be significant. If the difference is real, these results imply that the Cassiopeia source may be 10 or 20% smaller at the higher frequency.

Fringe Amplitude for various values of Base Line for a Wavelength of 10·7 cm

Base line (wavelengths)	Cassiopeia A	Taurus A
120	0·98 ± 0·05	0·94 ± 0·08
142	0·99	—
400	0·63	0·57
550	0·47	—
700	0·32	0·10

ACKNOWLEDGMENTS

The author would like to thank Professor A. C. B. Lovell, Mr. Hanbury Brown and other members of the Jodrell Bank staff for their helpful advice and assistance throughout the experiment. He is also indebted to the Department of Scientific and Industrial Research for a maintenance grant.

REFERENCES

HADDOCK, F. T., MAYER, C. H., and SLOANAKER, R. M., 1954, *Astrophys. J.*, **119**, 456.

HANBURY BROWN, R., JENNISON, R. C., and DAS GUPTA, M. K., 1952, *Nature, Lond.*, **170**, 1061.

HANBURY BROWN, R., PALMER, H. P., and THOMPSON, A. R., 1955, *Phil. Mag.*, **46**, 857.

JENNISON, R. C., and DAS GUPTA, M. K., 1956, *Phil. Mag.*, (Ser. 8), **1**, 55.

MILLS, B. Y., 1952, *Nature, Lond.*, **170**, 1063; 1953, *Aust. J. Phys.*, **6**, 452.

SMITH, F. G., 1952, *Nature, Lond.*, **170**, 1065.

LETTERS TO THE EDITOR

Current Noise in Indium Antimonide

An electrical network in thermal equilibrium exhibits noise the magnitude of which is expressed by Nyquist's theorem. The noise developed in a semiconductor carrying a direct current is usually very much greater than the thermal noise, and this often presents a severe limitation on the use of semiconductors in applications. A number of interesting devices using indium antimonide have been described by Ross and Saker (1955), and in consequence it was decided to investigate noise in this material. No evidence of such noise has, however, been obtained. When compared with the previous work on other materials this is an unexpected result. Although it does not seem possible at present to give a theoretical explanation, the experimental result is considered to be of sufficient interest to merit an early report.

Experiments have been carried out with single-crystal filaments formed by sandblasting into 'dumbell' shapes (Montgomery 1952) to produce current and potential contacts of large area. The filaments were approximately 1 mm × 1 mm in cross section and between 1·0 cm and 1·5 cm in length. The specimens were lightly etched, the contacts silver-plated and soft soldered into the measuring circuit. Low-resistance ohmic contacts can readily be made to indium antimonide in this way. Since the filament resistances were of the order of 1 Ω a specially constructed input transformer was required to obtain a satisfactory noise figure. The remainder of the apparatus consisted of a battery-powered preamplifier, Muirhead–Pametrada wave analyser (frequency range 20 c/s–20 kc/s) and electronic integrator. From the known noise figure of the apparatus and the minimum detectable change in noise power it is estimated that the noise ratio n (ratio of total noise power to Johnson noise power) was $n<7$ at 20c/s and $n<2$ for frequencies above 1 kc/s, for all filament currents up to 500 mA.

In germanium the measured noise ratio is consistent with a law of the form

$$n=1+AI^2/f+BI^2\{1+(f/f_1)^2\}^{-1} \qquad \ldots\ldots(1)$$

where I is the current, f the frequency and A, B, f_1, are constants for a particular specimen (van der Ziel 1954).

The third term describes a contribution which is designated 'shot noise' by analogy with the somewhat similar effect in vacuum tubes. This arises from the intrinsic fluctuation in the number of mobile carriers in the bulk of the material ($1/2\pi f_1=\tau$, the lifetime for bulk recombination). The expected contribution from this cause to current noise in InSb can be calculated for an intrinsic sample from the relation

$$n=\frac{e}{kT}\left(\frac{V_0}{d}\right)^2(\mu_n+\mu_p)\frac{\tau}{1+4\pi^2f^2\tau^2} \qquad \ldots\ldots(2)$$

where V_0 is the voltage across the noise-measuring contacts separated by a distance d; the electron and hole mobilities are μ_n, μ_p, respectively and f is the frequency. Inserting typical values; $V_0=1$ volt, $d=1$ cm, $\mu_n+\mu_p=6\times10^4$ cm^2 v^{-1} sec^{-1}, $\tau=10^{-6}$ sec, n is $1\cdot5\times10^{-3}$ for $4\pi^2f^2\tau^2\ll1$. This would be far too small to measure with our apparatus.

In germanium the second term in (1) is by far the most important for frequencies less than about 10^6 c/s. In a typical case (see van der Ziel 1954), $n = 10^3$ at 1 kc/s for $I = 3$ mA, and increases as the frequency is reduced down to very low frequencies (Rollin and Templeton 1953). Unfortunately there is no comprehensive theory of $1/f$ noise in semiconductors and although considerable progress has recently been made in showing how this type of noise can arise in germanium (McWhorter 1955, Bess 1956) it does not seem possible to extend these ideas to indium antimonide at present.

Acknowledgment is made to the Admiralty for permission to publish this letter.

Services Electronics Research Laboratory, D. J. OLIVER.
Baldock,
Herts.
14th December, 1956.

BESS, L., 1956, *Phys. Rev.*, **103**, 72.
MCWHORTER, A., 1955, *Lincoln Laboratory Tech. Rep.*, No. 80.
MONTGOMERY, H. C., 1952, *Bell System Tech. J.*, **31**, 950.
ROLLIN, B. V., and TEMPLETON, I. M., 1953, *Proc. Phys. Soc.* B, **66**, 259.
ROSS, I. M., and SAKER, E. W., 1955, *J. Electronics*, **1**, 223.
VAN DER ZIEL, A., 1954, *Noise* (New York : Prentice Hall), p. 206.

Note added in proof.—A similar result has recently been reported by G. H. Suits, W. D. Schmitz and R. W. Terhune, 1956, *J. Appl. Phys.*, **27**, 1385.

REVIEWS OF BOOKS

Numerical Analysis, by Z. Kopal. Pp. xiv + 556. (London: Chapman and Hall, 1955.) 63*s*.

With the development of the subject of numerical analysis, and with the prospects of considerable further developments in view of the particular facilities and limitations of automatic calculating machines, there has come to be room for two kinds of books on the subject, one concerned with the algebraical and analytical theory which may lie behind practical procedures for carrying out numerical calculations, either by hand or with an automatic machine, and the other concerned with those procedures themselves. This book is of the first kind, and its main value, which is considerable, will be to those concerned with the development of practical numerical procedures rather than to the users of those procedures; it will not be of much help to the physicist or engineer who has a specific numerical problem and wants some practical advice, as brief as possible, on how to set about finding a solution. Nor is it a suitable book for a beginner: it is too comprehensive and too general. For example, the beginner is likely to be confused rather than helped by being presented with a whole array of interpolation formulae, those of Newton, Gauss, Stirling, Bessel, Everett, Steffensen, Chebycheff–Bessel, Chebycheff–Everett and the modified forms of the two last, all with their algebraic derivations; what he wants is one or two formulae which are easy to use in practice and will give the results he wants; and, if he is a physicist or engineer, these results are likely to be numbers of five figures or fewer, rather than numbers of ten figures or more which may be the main field for some of the more sophisticated formulae.

Some of the treatment seems rather unrealistic from the point of view of the reader who really does want to carry out a practical numerical calculation. For example, concerning the Picard method for integration of a differential equation the author gives a formal proof, and emphasizes by printing in italics, that " the maximum error of the nth approximation can be made less than any preassigned quantity provided only that n has been taken large enough ". To anyone who has really considered what would be involved in evaluating, say, e^{-50} by Picard's method applied as a numerical process to the equation $y' = -y$, the word ' only ' here might seem to be meant ironically! The emphasis here, and similar wording elsewhere, suggests the outlook of the pure mathematician rather than an appreciation of the realities of practical numerical work. To generalize a statement of Jeffreys, " in order that a series should be useful for numerical computation, it is neither necessary nor sufficient that it should converge "; and this should never be far out of mind for a writer on numerical analysis. Another example of unreality is the dismissal of the subject of integration of a first-order linear equation by a footnote stating that such an equation " would be integrable in terms of quadratures "; so it would, but it does not follow that this would be the best practical way of dealing numerically with such an equation. However, the possibility that other methods may be needed in practice is not even mentioned.

For a book of this size there are some surprising omissions besides that just mentioned. Almost two whole chapters—about 100 pages—are devoted to boundary-value problems, in the restricted sense of characteristic-value problems of homogeneous linear differential equations with homogeneous boundary conditions. However, other problems of differential equations with 'jury' conditions (to use L. F. Richardson's term) are dismissed in a single sentence, " non-linear systems would call for an essentially different method of approach "; so they would, indeed, but this bare statement is not much help to a reader faced with such a problem and needing some advice about how to deal with it. And even for linear equations, there is no mention of the method, which may be more useful in practice than any mentioned in the text, of integrating from both ends of the range and matching at some intermediate value of x. Nor would the reader get any hint of the scope of the relaxation process applied to ordinary differential equations with two-point boundary conditions, or of the application in this context of Choleski's matrix factorization procedure, as developed by Thomas and by Fox. Some of the detailed algebra of Chapter V could well be spared to give space for consideration of these procedures.

However, the reader with some previous knowledge of the subject, and with practical experience of the realities of numerical work, will find much of value in this book. Topics which may be selected as being of particular interest are the optimum size of interval in numerical differentiation, a thorough treatment of quadrature formulae of Gauss type, the use of Chebycheff polynomials in approximate representation of functions by polynomials and in interpolation, an original study of methods for replacing a differential equation by a difference equation of higher order without introducing a dominant spurious ' solution ', and the numerical treatment of integral and integro-differential equations. And not the least merit of the book is that it provokes thought and discussion on many of the topics with which it is concerned. D. R. HARTREE.

Solid State Physics, edited by F. SEITZ and D. TURNBULL. Advances in Research and Applications, Vol. 1. Pp. xii + 469. (New York: Academic Press; London: Academic Books, 1955.) $10.00.

This volume is the first of a series to be written on the Physics of the Solid State under the editorial direction of Professor Seitz and Dr. Turnbull. The purpose of the series is to provide up-to-date information covering all important aspects of solid state science. The first two sentences of the preface are worth quoting: " The knowledge of solid state physics has grown tremendously since 1940. This growth has occurred on an international scale and perhaps has been more rapid, in relation to the 1940 effort, than that in most other major fields of physics." When the enormous advances in nuclear physics and the spectacular discoveries of fundamental particles are recalled this quotation shows the degree of importance which the editors attach to their subject. The tremendous scope of even a single field in modern physics appears to be changing the means employed for the dissemination of knowledge in this science. The book by a single author appears to be giving way to the composite volume with chapters written by different contributors, each expert in his own branch. There is inevitably a loss of unity of style and thought in the presentation, but against this there is a gain in the speed with which new results can be made readily accessible, and probably also a gain in the extent of the detailed information which can be made available.

The present volume contains six articles, five of which are concerned with one or other aspect of the electron theory of solids; the remaining article deals with the theory of order–disorder transformations in alloys. The first is by J. R. Reitz and is entitled Methods of the One-Electron Theory of Solids. This deals in a thoroughly competent manner with the methods of determining electronic wave functions and eigenvalues in crystals. The various methods now well established are reviewed: the cellular method, the variational method and the method of orthogonal plane waves. The nature and properties of Brillouin zones are described, and towards the end there is a short account of the Heitler–London method as applied to ionic crystals. The second article, by a famous pair of authors, Wigner and Seitz, is a qualitative analysis of cohesion in metals. This is the shortest of the six articles and is substantially the same as the contribution made by Wigner at the International Conference on Theoretical Physics held in Tokyo in 1953. It contains a useful table of experimentally determined cohesive energies of monatomic solids. F. S. Ham contributes an article on the Quantum Defect method of calculating energy bands in crystals. This method originated with the work of Kuhn and Van Vleck and is particularly well suited to the alkali metals. An interesting account, in some 85 pages, of valence semiconductors is given by H. Y. Fan of Purdue University. This chapter contains a great deal of information and will be found useful to a variety of readers. It succeeds perfectly in summarizing the state of knowledge of the bulk properties of germanium and silicon up to 1955. However, the reader wishing to obtain a complete grasp will find himself compelled to have recourse to the original literature on many occasions. To help him in this way a very complete set of references is provided.

The last article of the book is by David Pines and deals with the interaction of electrons in metals. This theory of the plasma oscillations in metals represents probably the most important single advance in electron theory in the past decade, and the account given here will be of interest to many readers. Not only is the basis of the theory made clear in a simplified way, but many applications are given which illustrate the significance of the new ideas.

In a short review it is impossible even to mention all the topics dealt with in this volume. It must suffice to say that it can be thoroughly recommended to all interested in solid state physics. It is a beautifully produced book; the print of both words and mathematics is clear and easy to read, and perhaps, not the least of its merits is its convenient size; it is appreciably smaller than most volumes of its kind. H. JONES.

The Mechanical Properties of Textile Fibres, edited by R. MEREDITH. Pp. xii+333. (Amsterdam: North-Holland Publishing Co., 1956.) 61*s*.

This book is a welcome contribution to the literature of an important class of high polymers—namely the natural and artificial textile fibres. Although different parts of the book are written by different authors, they have cooperated extremely well and the book has a unity often lacking in such cases.

The elastic behaviour of most textile fibres is very complicated and the field covered by the authors is very wide. In the present state of knowledge no one approach can elucidate the behaviour of textile fibres in extension. In the introductory chapters Drs. Peters and Woods have considered fibres in general as chain molecular systems and outlined the possible mechanisms involved in the extension of such systems. More than one mechanism is frequently involved. The authors have also indicated the various factors involved, such as degree of crystallinity, cross linking, orientation and chain folding. Thermodynamic analysis of experimental results can be used to determine whether extension involves an increase in internal energy or a decrease in entropy.

These considerations have been amplified in the remainder of the book which deals with the properties of different fibre types. This is divided into three sections dealing respectively with the natural and artificial cellulose fibres, the natural and regenerated protein fibres, and the synthetic fibres. The section on protein fibres is the work of Drs. Peters and Woods and the editor has been responsible for the other two sections. Each section begins with a chapter on the structure of the different fibres, and is followed by a chapter on the time effects found in stretching fibres.

The stress–strain curves are discussed experimentally and theoretically in subsequent chapters. In discussing the theory of initial Hookean extension of keratin fibres, Drs. Peters and Woods make use of hitherto unpublished work based on the Morse function for inter-atomic forces. In so doing they give a table of the elastic moduli to be expected from different types of bond. There appears to be an error in this table in that the hydrogen bond energy is taken as less than that for van der Waals binding. The latter is usually taken as the weakest type of inter-atomic or inter-molecular bonding, and therefore should have the lower energy. The general argument is not seriously affected as the uncertainty in the number of bonds per unit chain length, to which the authors draw attention, makes any exact comparison between calculated and observed moduli impossible in the present state of knowledge.

This criticism raises the question as to how far the inclusion of unpublished theories is justified in a book of this type. In the normal process of publication in a scientific journal errors of the above type would probably be detected during refereeing and eliminated.

Chapters are also devoted to the various theories of visco-elastic behaviour of which the most important are the reaction-rate theories and dynamical theories based on spectra of relaxation times. In the section on protein fibres one chapter is devoted to setting phenomena, which are so important with wool in particular.

All these aspects of the tensile behaviour of fibres have been dealt with very comprehensively. Much less attention has been given to the torsional behaviour of fibres for the simple reason that research workers generally have been concerned mainly with tensile properties. One source of difficulty is the shape of the cross section of many textile fibres which makes it difficult to relate torsion to fundamental elastic properties. The authors appear to have assumed (p. 86) that the torsional rigidity is solely determined by the rigidity modulus and the polar moment of the cross section. This is not true for most non-circular cross sections as has been pointed out by Lamb (*Statics*) and Timoshenko (*Theory of Elasticity*, 1934, p. 228) the shape factor being in general less than that calculated on the above assumption. Reference to original papers (by the editor) suggests that the shape factors quoted may have been calculated from more exact relations. In a book of this standard it should have been clearly stated that the equation given for the shape factor is only approximate.

Of minor misprints the ratio quoted in Table 4 p. 122, might be pointed out although it is obvious.

These criticisms are not meant to detract from the book which is one that everyone interested in the physical properties of high polymers generally and of fibres in particular, cannot afford to be without. It should be in every textile physics laboratory, and for a book of this standard the price is a reasonable one.

F. L. WARBURTON.

Champs de vecteurs et de tenseurs. Introduction à l'électromagnétisme, by E. BAUER. Pp. vii + 204. (Paris: Masson et Cie, 1955.) 2200 *fr*.

In the preface to this beautifully written account of the vector and tensor calculus, the author says: "J'ai toujours pensé que, pour bien comprendre la physique, pour distinguer nettement les lois données par l'experience, ou les hypothèses théoriques, de ce qui est simple identité mathématique, il est nécessaire de préciser d'abord le langage mathématique, d'en faire un cadre abstrait, général, dans lequel pourra venir s'insérer ensuite notre connaissance des faits. Une fois achevé ce travail préliminaire, le terrain est déblayé, les réalités physiques fondamentales se dégagent de la gangue de calculs dans laquelle elles sont trop souvent noyées."

Professor Bauer has succeeded admirably in his aim of showing that an elegant abstract mathematical scheme provides the natural setting for the description of the physical world. His book includes a detailed account of the theory of vector and tensor fields in three dimensions, a chapter on general tensor theory and an excellent brief modern introduction to electromagnetic theory. The lucid simple style of the book should commend it particularly to readers (such as the reviewer) whose command of the French language is limited, and its treatment of the subject should serve as an example both to those who still regard 'applied mathematics' primarily as a parade ground for the exercise of pure mathematical tricks, and to those extremists of the opposite camp (if any still exist) who remain to be convinced of the need for precise mathematical language in the description of nature.

E. H. SONDHEIMER.

Ionized Gases, by A. VON ENGEL. Pp. x + 281. (Oxford: Clarendon Press; London: Cumberlege, 1955). 42*s*.

In recent years the subject of discharge in gases has been of such wide interest that several very bulky books on it have been published, each giving extensive references and being as up to date as possible at the date of publication. Such books can bemuse the less experienced reader by their mass of detail, and it is not always clear to him whether statements are to be taken as established fact, the opinion of the original author, or a comment by the book author. The appearance of the present book, a relatively slim volume, by one of the authors of a well-known two-volume work published some twenty years ago, is sure to attract notice from the less highly specialized reader. In his preface, the author states that the text is based on lectures to undergraduates in physics, and that it " puts emphasis on the fundamental concepts and the limitations of the various treatments ".

The first three chapters give some historical notes, conduction in feebly ionized gases, and the production of charged particles. This is good introductory material, but the arrangement is such that, in a number of cases, the reader is asked to accept statements that will be discussed in later chapters. There follow chapters on mobility, diffusion, and recombination which are a useful survey of present opinions. But the last three chapters on ionization in an electric field, glow discharge, and arc discharge will be of principal interest to most who turn to the book—those seeking a background for work in applied science. The spark discharge itself is dealt with under ionization in the electric field. In this latter section, the treatment could have been made more objective in style. The possible secondary processes receive very brief mention, and attention is quickly devoted to that due to positive ion bombardment at the cathode to continue the development of the text, and the occasion is missed of presenting the concept of a generalized secondary coefficient. The few lines given to the sphere gap (p. 175) are quite inadequate, the curvature of the breakdown voltage vs. spacing characteristic being dismissed as field distortion such that " the field along the line of centres is larger than V_s/d ", and the effect of asymmetry is merely indicated as a lowering of V_s for $d>r$, but without explanation or discussion. The streamer theory receives a paragraph on p. 186, but leaves one with the impression that it was satisfactory for all but very large gaps. In the remarks on formative time-lag, the importance of the linear relation between time-lag and spacing for a given percentage overvoltage applied is not mentioned.

This should be a very useful course book for the physics student, but it requires to be supplemented by other reading if he is to enter the field of applied physics, in which the complexities are much greater than the text suggests. For the same reason, it will be of more limited interest to researchers or engineers. It is produced in the well-established presentation and finish of the Clarendon Press. F. M. BRUCE.

Techniques générales du laboratoire de physique, by J. SURUGUE. 2nd Edn. Pp. 671. (Paris: Centre National de Recherche Scientifique, 1955.) 2400 *fr*.

This is an important member of the little class of books which deal with laboratory technique—it is a revised and enlarged edition of a book published in 1947 and it is to be followed by at least one more volume. Such a book can hardly avoid being eclectic and a good deal of its value lies in providing examples and ideas which are different from those in other books at the disposal of the reader. It is therefore profitable to give a critical survey of its chapters. The first chapter deals with mechanical design, and is not particularly distinctive, though its examples and details are valuable; it is good on the uses and design of castings. The section on glassblowing is written to help the elementary learner of craftmanship rather than the physicist, who in designing his experiments might need to know about the potentialities of the glassblowing lathe, the properties of special glasses, or the principles of metal–glass seals.

A section on vacuum technique introduces the ideas of low-pressure flow, and describes a number of pumps and gauges. Although it deals with many devices in current use, some of its paragraphs have a slightly old-fashioned look—for example, it describes a falling-drop mercury pump of Sprengel type but does not deal with the important fractionation principle applied to oil diffusion pumps; it discusses waxed joints and rubber-ring joints without emphasizing that the latter are far more useful than the former in the design of metal vacuum systems.

The chapter on furnaces and high temperature contains a deal of information not easily available elsewhere, and includes a specialized section on measuring the temperature of flames. The section on optical techniques gives the best résumé I know of the peculiarities of optical materials, the choice of optical components and the adjustment of optical systems. It is followed by a section on light sources and filters. The section contains a number of items of electrical and vacuum technique which have a wider application. Next comes an authoritative, up-to-date, but too brief review of photoelectric devices. A chapter on methods of recording reviews a field better known to the industrial and military instrument engineer than to the laboratory physicist, and it is therefore of considerable interest to the latter: the principal adverse criticism is that few hints are given, explicitly or by reference, about what is actually available from the instrument manufacturer.

A chapter on rectifiers and on voltage and current stabilizers gives a useful and quantitative introduction to the subject, but the stabilizers described do not represent the present state of the art. A long chapter, new since the 1947 edition, deals with electronic circuit techniques. It begins with the Fourier analysis of a number of waveforms which might appear in the physical laboratory, and discusses the way in which they are treated by amplifying circuits; it does not stress the simple and revealing direct investigation of the progress of voltage steps through such circuits. The section contains a number of rather skeletonized circuit diagrams, drawn on conventions which make them rather trying for the English reader. It is clear that the book is one which should be available for reference by research students, and that its usefulness to the individual worker depends very much on his special interests. H. J. J. BRADDICK.

Abstracts of the Literature on Semiconducting and Luminescent Materials and their Applications (1954 Issue), compiled by the BATTELLE MEMORIAL INSTITUTE. Pp. viii + 200. (New York: John Wiley; London: Chapman and Hall, 1955.) 40*s*.

This, the second volume of abstracts compiled by the Battelle Memorial Institute, will be of great value to engineers and scientists working in the field of semiconductors. The present volume covers papers published in 1954, together with a few foreign (i.e. not American) publications in the period September 1953 to 1954 which were not included in the 1953 issue.

There are 750 abstracts in the Issue classified as before, on the basis of materials. About a third of the abstracts are of papers on germanium and silicon but organic semiconductors, oxides, sulphides and halides receive very adequate coverage. There are subject and author indexes.

The emormous amount of information published annually in the field of semiconductors presents a formidable problem to anyone actively engaged in the field. If he is to have time to make any significant contribution himself he is unlikely to have time to read some 750 papers a year, published in some dozens of journals all over the world. Abstracts of the literature, provided they are published with the minimum of delay, are invaluable. The Electrochemical Society and the Battelle Memorial Institute deserve the thanks of all in the semiconductor field for undertaking this project. The first (1953) Issue had a number of noticeable defects, the most serious of which was inaccurate abstracting. The present volume is greatly improved in this and other respects. Indeed, your reviewer has been unable to find any errors in the abstracts.

It would seem desirable, if this book is to be issued annually for many years to come, to improve the binding. At present the sheets are held, as in a 'loose leaf' book, by a helical wire and the cover is of stiff paper. This seems inadequate if the book is to be used for frequent reference in the average laboratory library. If the adoption of a more conventional binding would materially slow up publication perhaps at least a stiff cardboard cover could be used.

A. F. GIBSON.

Atoms and the Universe, by G. O. JONES, J. ROTBLAT and G. J. WHITROW. Pp. 254. (London: Eyre and Spottiswoode, 1956.) 25*s*.

This is an interesting but strangely uneven book. It deals mainly with recent advances in the fields of nuclear physics and of astrophysics—with the almost unimaginably small and almost unimaginably vast. These are the two extremes of physics concerned with the behaviour of matter in conditions far removed from those of every-day events in our domestic physical world, yet they have much in common. The concept of bringing them together in one volume is original and stimulating.

The first five chapters are on Elementary particles, The atom breaks up, Smashing the atom, Energy from the atom and Cosmic radiation. The last four chapters deal with The solar system, The Milky Way, The size of the universe and The age of the universe. Both these sections are admirably up-to-date accounts of the kind one would hope for in an article in a first-rate encyclopaedia. They are sufficiently simple to be understood by the intelligent layman; they are sufficiently comprehensive to be read with interest and profit by the physicist who has not specialized in nuclear physics or astrophysics.

But one cannot help feeling that the book would have been better balanced if the authors had been content to restrict it to these two sections. The introduction, the chapters on Matter and its properties and From classical to modern physics, and the Appendix are scrappy and breathless in their haste. An attempt had been made to compress far too much into too small a space. One is reminded of the occasion when a lecturer realizes that he has only five minutes left of the appointed hour and finding he has not yet covered half of the matter which he had planned, attempts feverishly to race through the rest. The plan of the book is too ambitious; the nucleus and the universe are in themselves very considerable subjects. To try to include as well an account of the relation of physics to chemistry and biology, the methods of experimental science, relativity, the structure of matter, the quantum theory, a disquisition on theories and the role of mathematics, and laws, hypothesis and theories—no, the reader rebels. One cannot compress an account of the properties of matter into 24 pages, or of the philosophy of science into a few paragraphs and it would have been more artistic to refrain from the attempt.

In spite of these blemishes, it is a very readable book. It is hard to get a good general view of the various branches of science in these days of intense specialization, and one is grateful for this excellent summary of what the nuclear physicists and astronomers are discovering and thinking.

W. L. BRAGG.

Small-angle Scattering of X-rays, by A. GUINIER and G. FOURNET. Pp. xii + 268. (New York: J. Wiley; London: Chapman and Hall, 1955.) £3. (English translation by C. B. Walker.)

This very comprehensive book contains a detailed survey of the basic theory of x-ray scattering (most of which is equally applicable to light scattering) occupying about one-third of the whole volume; the remainder of the book is about equally divided between experimental equipment, interpretation of experimental data, present-day applications, and a very extensive bibliography.

The theory of low-angle scattering is developed, first for a single particle, then for a dilute suspension of particles for which the particles may be considered to scatter independently, and finally for close-packed systems in which the interaction between particles must be taken into account. Any mathematical exposition of this kind is liable to be attacked from two sides; by the kind of rigorist who wants a full and generalized theory with nothing left out, on the one hand; and by the sort of experimentalist who wants only a useful compendium of formulae, on the other. The treatment given here is perhaps more likely to satisfy the rigorist, and this is good, not only because it fills a present need for a complete exposition of the theory, but also because most of the formulae used in the interpretation of experimental data are not unconditionally valid relations, but are mostly approximations to the truth. It is most important that users of these formulae should thoroughly understand the nature of these approximations and the range of their validity, and this chapter should help them to do so. However, the reviewer cannot help feeling that there are one or two places where the theory is unnecessarily frightening; for instance, as early as p. 8 there is a subsection (2.1.2.1) which is enough to scare many an experimentalist away; yet this part is apparently not essential to the main argument until much later.

The chapter on experimental methods and equipment includes a full treatment of the governing conditions relating x-ray beam width and intensity with exposure time and resolving power for long periodicities. The various systems of plane and focusing monochromators are fully described, with their merits and drawbacks.

The chapter on interpretation shows, with the aid of examples from the literature, what conclusions can be drawn from the experimental data. A short chapter follows on the comparison of the results of x-ray low-angle scattering with those of electron microscopy and the measurement of line broadening in high-angle x-ray diffraction photographs.

Among the applications described are investigations on solutions, soaps, biological fibres, catalysts, and alloys. A much-needed warning is given of the dangers of drawing too sweeping conclusions from inadequate data, especially when these conclusions may be used by workers in other fields who do not fully understand the limitations inherent in x-ray diffraction methods. One might well quote (p. 175), " It is necessary to retain only those results that have been obtained by a restrained application of a correct theory to the [diffraction] patterns, leading to a modest but sure interpretation ". This is typical of the scientific integrity which pervades the whole book.

The bibliography, compiled by K. L. Yudowitch, contains 569 titles, arranged by subjects in seven sections. With many of these a very brief abstract is given. There are also a few references at the end of each chapter. The advantage of this double system of references is open to question.

The translation is good, in that the reader is not generally conscious that the original was not written in English. The book is well produced, and the price reasonable considering the quantity of useful material given.

A. R. STOKES.

Wind Tunnel Technique, by R. C. PANKHURST and D. W. HOLDER. Pp. xviii + 702. (London: Pitman, 1952). 70*s*.

From its early days, a little before the first world war, wind tunnel technique has developed and expanded tremendously, the wind speeds involved having risen by a factor of about fifty, and the powers by roughly a thousand. As a result of the work done, many thousands of reports have been issued and it is indeed surprising that we have had to wait so long for this, the first comprehensive account to be published in this country of the tools and methods employed. Let me hasten to add that the waiting has not been in vain, for Pankhurst and Holder have given us a fine book which should be included in the library of every wind tunnel department. Their expressed aims of giving a critical résumé for graduates entering the field of experimental aerodynamics, and of providing a work of reference for the staffs of wind tunnels in research establishments and in the aircraft industry seem to me to have been amply fulfilled.

After an introduction which includes a summary of the elementary equations of compressible flow, there follows an excellent chapter on the design of wind tunnels in which the authors discuss power requirements, aerodynamic design for flow uniformity and low turbulence, and give details of many types of wind tunnel. A chapter on flow visualization ranges from the use of smoke and silk tufts to schlieren, shadowgraph and interferometer methods. Next come four chapters dealing with the commonest quantitative measurements involved in wind tunnel work, and the apparatus used for these measurements; these cover pressures, velocity and flow direction, forces and moments on models, manometers and balances. Interference effects, tunnel constraint corrections, reduction of observations, and some suggestions on presentation of results form the subject of two chapters. In a Special Measurements chapter, the sections of necessity somewhat brief include boundary layer, turbulence, stability and control, flutter, and propellers. Here I would have preferred less space devoted to propeller tests and more to the representation of jet flow in wind tunnel models and I was sorry to find no mention of jettisoning tests. The penultimate chapter is devoted to experimental analogies including the electric tank, the Hele–Shaw analogy for two-dimensional steady potential flow, and the surface-wave analogy for two-dimensional supersonic flow, and the last chapter to models, model strength and rigging. Finally, four appendices summarize a great deal of frequently needed data and miscellaneous information not readily found elsewhere; amongst this, I was glad to find aerofoil notations and foreign aerodynamic symbols.

The book is well presented, its 373 illustrations are admirably reproduced, and a great feature is that some six hundred excellent references are given, grouped at the end of appropriate chapters. In conclusion, I have already found many occasions for referring people to this book in response to requests for advice, and I am sure my copy of " Pankhurst and Holder " will become as well-thumbed as " Glauert " has been in the past.

A. SPENCE.

The Electrolytic and Chemical Polishing of Metals in Research and Industry, by W. J. McG. TEGART. Pp. ix + 129. (London: Pergamon Press Ltd., 1956.) 30*s*.

In an age when mechanization has increased to the point of automation, it is interesting to reflect that the polishing of metals, for so long a process of a mechanical nature, is for many purposes now replaced by much simpler and quicker electrolytic and chemical methods.

This small book surveys this field and the related subject of chemical polishing in an admirably concise and lucid manner, and the clear print and presentation characteristic of the printers and publishers add still further to the readers' enjoyment and ease of assimilation of the contents.

Chapter 1 goes to the root of the matter in outlining clearly the evidence of various kinds indicating the nature of the process of electrolytic polishing, as far as it has by now been ascertained. The characteristic relation between the anode current and voltage is first described ; then the general effect of the temperature, concentration, viscosity and stirring of the electrolyte are discussed. This is followed by the evidence showing that while often a viscous layer is formed in the electrolyte next to the anode, it is more essentially the formation of a surface film of oxide, chloride, cyanide or other compound (according to the nature of the bath) that promotes the submicroscopic smoothing or 'polishing' action instead of either the preferential etching or larger scale levelling of the surface. The use of this basic information to develop baths and conditions suitable for specific metals or alloys is shown, and finally the theories and special features of electropolishing in molten electrolytes are considered.

Chapter 2 details helpfully the electrical characteristics of the polishing process; chapter 3 briefly discusses factors influencing polishing conditions (surface preparation, temperature, time, agitation, etc.); chapters 4–7 deal with laboratory and research applications of electropolishing, general principles, precautions in use of perchloric acid and other reactive solutions, and an especially helpful table of laboratory methods for specific metals and their alloys. Chapters 8 and 9 discuss industrial applications generally and for specific metals and alloys; and 10 and 11 deal with chemical polishing and its applications in the laboratory, and industrial methods for specific metals and alloys. Finally, three appendices show rates of anodic dissolution, amounts of metal removed by mechanical and electrolytic polishing, and details of analysis of solutions; and there is a bibliography of 213 references, and an index. This handy little book will find a useful place in any laboratory and many fields of industry.

H. WILMAN.

CONTENTS OF SECTION A

PAGE

Dr. N. H. March. A Variational Method for the Calculation of Particle Densities and Sums of Eigenvalues in Wave Mechanics 169

Dr. A. Dalgarno and Mr. N. Lynn. Curve Crossing and Orthogonality of Molecular Orbitals 176

Dr. F. Mandl. Dispersion Relations for Forward Scattering of Spin Zero Particles by an Extended Potential 183

Dr. C. Henderson and Mr. A. Scott. The Multiple Scattering of Positrons and Electrons 188

Mr. D. L. Allan. Protons from the Interaction of 14 mev Neutrons with Medium Weight Nuclei 195

Mr. N. E. Booth, Mr. B. Ledley, Mr. D. Walker and Mr. D. H. White. Nuclear Cross Sections for 900 mev Protons 209

Dr. R. F. Barrow and Prof. E. Miescher. Fine Structure Analysis of NO Absorption Bands in the Schumann Region 219

Research Notes :

Dr. A. Dalgarno and Mr. N. Lynn. An Exact Calculation of Second Order Long Range Forces 223

Letters to the Editor :

Prof. N. Feather. Note on the Population of Segments of a Random Series which include the Origin 226

Reviews of Books 228

Contents of Section B 232

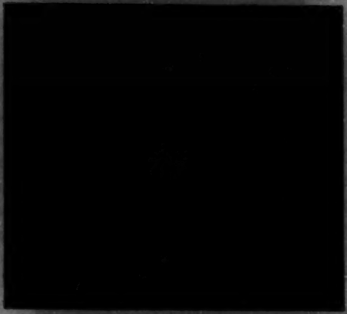


Figure 3 (*a*). Transmission pattern showing secondary diffraction spots.

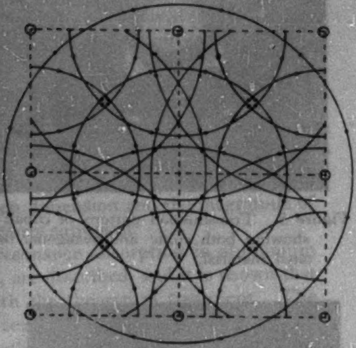


Figure 3 (*b*). Diagram showing secondary scattering of 200 and 220 Ni beams by NiO in (111) orientation.

Figure 4. Pattern from very thin Co layer on (001) NaCl.

Figure 5. Transmission pattern of cobalt film showing both cubic and hexagonal orientated material.

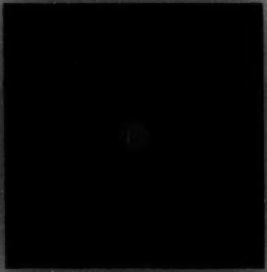


Figure 6. Transmission pattern of Fe film grown on (111) NaCl.

Figure 7. Pattern from iron specimen one minute after deposition.

Figure 8. A later stage in the oxidation.

Figure 9. The complete spinel structure.

The Effect of Eddy Currents on Nuclear Magnetic Resonance in Metals

By A. C. CHAPMAN, P. RHODES and E. F. W. SEYMOUR
Physics Department, University of Leeds

Communicated by E. C. Stoner; MS. received 30th October 1956, and in final form 8th December 1956

Abstract. A theoretical and experimental investigation of this effect is described. The theoretical treatment is developed for specimens in the form of (i) a flat plate, (ii) a long cylinder and (iii) a sphere; and is applicable to foils, wires and powders with spherical particles. In each case the power absorption per unit volume W is expressed as $W = W_0 + W_1$, where W_1 determines the nuclear resonance absorption and is of the form $W_1 \propto a\chi' + b\chi''$, where $(\chi' - i\chi'')$ is the complex susceptibility. Explicit expressions are derived for W_0, a and b as functions of $p = 2t/\delta$, where δ is the skin depth, and $2t$ is (i) the thickness of the plate, (ii) the diameter of the cylinder, (iii) the diameter of the sphere. These results indicate that both the dependence on reduced specimen dimension and the magnitude of the effect differ considerably from those previously supposed.

Measurements of the nuclear magnetic resonance absorption in aluminium foils, in the temperature range 20°K to 300°K, are described, and it is shown that the results are in satisfactory agreement with the theoretical treatment. The effect of eddy currents on the position of the centre, the shape and the amplitude of the absorption line are discussed, and methods of correcting observed lines to obtain the undistorted line are indicated. Extensions of the treatment to methods of observing nuclear resonance other than by absorption measurements are also outlined.

---

§ 1. Introduction

The observation of nuclear magnetic resonance in a conducting material is complicated by the fact that the radio-frequency magnetic field used to excite resonance sets up eddy currents in the material. These currents effectively confine the magnetic flux to a surface layer of thickness of the order of the 'skin depth'. An obvious result is that the intensity of the nuclear absorption per unit volume of material is less for large particles than for small ones. A more subtle effect arises from the resonant change in flux distribution, and hence in eddy current loss. This change in eddy current loss may be observed with the nuclear absorption, with the result that the resonance line shape is affected. Bloembergen (1952) has shown that the observable absorption line shape is then determined by both the real and imaginary parts of the complex nuclear susceptibility,

$$\chi = \chi' - i\chi'', \qquad \ldots\ldots(1.1)$$

and has given explicit formulae applicable to several limiting cases. For particles with the relevant dimension small compared with the skin depth, conditions revert to the simple situation typical of non-metallic specimens, in which the line shape is determined by χ'' alone. In general, however, both χ' and χ'' are involved, and there is a distortion and shift of the absorption line. The form of the dispersion line is also modified.

In the course of some investigations at low temperatures, involving comparisons of nuclear magnetic resonances in pure metals with those in their alloys, it became clear that it would be difficult to realise the desirable condition of particle dimensions small compared with the skin depth, which would probably be of the order of 10^{-3} mm for the purest samples. Further, since the electrical conductivities of the alloys would vary with solute concentration, the distortions arising from the eddy currents would also vary. Corrections for this effect would thus be required before significant comparisons between the alloys could be made.

A theoretical treatment has, therefore, been developed, in extension of that of Bloembergen (1952), to determine the effect of specimen shape and size on nuclear resonance absorption, and hence to provide a means of deducing the characteristics (e.g. intensity, shape and location of the centre) of true nuclear absorption lines (corresponding to χ'') from observed lines. This treatment leads to some somewhat unexpected results, and an experimental investigation has been made to verify these.

The theoretical treatment and the results obtained are given in § 2, and the experimental investigation is described in § 3. Some implications of the results of the present work, and the extension of the treatment to methods of observing nuclear resonance other than by absorption measurements, are discussed in § 4.

§ 2. THEORETICAL

There are several methods which may be used to calculate the power absorption in a conducting body in an applied alternating magnetic field. The choice of method depends in part on the shape of the body considered. In most nuclear resonance experiments on metals the samples are in the form of either foils, wires or powders. It was considered appropriate, therefore, to develop the theoretical treatment for specimens in the form of (i) an infinite flat plate, the results here being applicable to a foil in which the thickness is small compared with the other dimensions; (ii) a cylinder of infinite length, the results being applicable to a wire with diameter small compared with the length; (iii) a sphere, the results being applicable to powders with spherical, or nearly spherical, particles. For the plate and cylinder the most direct method of calculation is to use the complex Poynting vector to find the time average of the power flow into the body. For the sphere a method due to Casimir (1947), who derived an expression for the magnetic moment of a conducting sphere in an alternating field, is more convenient.

By these methods it is possible to derive general expressions for the power absorption in a plate, cylinder and sphere (cf. equations (2.4), (2.17), (2.25)). These expressions, as obtained, are not, however, in a form suitable for immediate application, since the requirement is for the explicit dependence of power absorption on χ' and χ''. For the plate and sphere, Bloembergen (1952) has derived approximations to this dependence appropriate in the limits of very small and very large specimens. It is shown below, however, that under the conditions normally satisfied in nuclear resonance, i.e. $\chi' \ll 1$, $\chi'' \ll 1$, it is possible, for all three shapes of specimen, to derive explicit expressions, in closed form, for the power absorption. In each case the power absorption per unit volume W may be expressed as

$$W = W_0 + W_1 \qquad \ldots\ldots(2.1)$$

where W_0 is independent of χ' and χ'', and corresponds to the 'ordinary', off-resonance, eddy current absorption; and W_1 is proportional to a linear combination of χ' and χ''.

(i) *Plate.*

For a flat plate of thickness $2t$ in an alternating magnetic field of amplitude H_1 and angular frequency ω parallel to the surface, it may be shown from Maxwell's equations and the expression for the Poynting vector that the power flow per unit area P into each face of the plate is given (cf. Bloembergen 1952) by

$$\begin{aligned} P &= \tfrac{1}{2}\mathscr{R}\left\{\frac{c}{4\pi}\mathbf{E}\wedge\mathbf{H}^*\right\} \\ &= \tfrac{1}{2}\mathscr{R}\left\{-\frac{i\omega\mu H_1^2}{4\pi K}\tanh(Kt)\right\} \qquad \ldots\ldots(2.2) \end{aligned}$$

where

$$K^2 = 4\pi i\omega\mu\sigma/c^2. \qquad \ldots\ldots(2.3)$$

In these equations $\mathscr{R}$ denotes the real part of the complex quantity, H^* is the complex conjugate of H, μ and σ are the permeability and conductivity of the material of the plate, and Gaussian units are used. In (2.3) a term involving the displacement current has been neglected since it has no appreciable effect in the present application. The total power absorption per unit volume W is given by

$$W = -\frac{\omega H_1^2}{8\pi t}\mathscr{I}\left\{\frac{\mu}{K}\tanh(Kt)\right\} \qquad \ldots\ldots(2.4)$$

by where $\mathscr{I}$ denotes the imaginary part of the complex quantity. It is convenient to introduce a skin depth δ corresponding to unit permeability and defined by

$$\delta = c/(2\pi\omega\sigma)^{1/2}. \qquad \ldots\ldots(2.5)$$

Also, the permeability μ is complex and may be written (cf. (1.1))

$$\mu = 1 + 4\pi(\chi' - i\chi'') = 1 + 2(\alpha - i\beta) \qquad \ldots\ldots(2.6)$$

where $\alpha = 2\pi\chi'$, $\beta = 2\pi\chi''$. It is now required to obtain from (2.4) an explicit expression for W as a function of t, χ' and χ''.

From (2.3), (2.5), (2.6), and noting that $\alpha \ll 1$, $\beta \ll 1$, it may be shown that

$$(\mu/K)\tanh(Kt) \simeq X_0 + X_1 \qquad \ldots\ldots(2.7)$$

where

$$X_0 = \tfrac{1}{2}(1-i)\psi\delta \qquad \ldots\ldots(2.8)$$

$$X_1 = \tfrac{1}{2}(1-i)\{\psi + (1-\psi^2)z\}(\alpha - i\beta)\delta \qquad \ldots\ldots(2.9)$$

$$z = (1+i)t/\delta, \qquad \psi = \tanh z. \qquad \ldots\ldots(2.10)$$

Then, from equations (2.3) to (2.10), W may be expressed in the form (2.1), with

$$\frac{8\pi W_0}{\omega H_1^2} = \frac{1}{p}\left\{\frac{\sinh p - \sin p}{\cosh p + \cos p}\right\} \qquad \ldots\ldots(2.11)$$

$$\frac{2W_1}{\omega H_1^2} = a\chi' + b\chi'' \qquad \ldots\ldots(2.12)$$

where

$$a = \frac{\sinh p - \sin p}{2p(\cosh p + \cos p)} + \frac{\sinh p \sin p}{(\cosh p + \cos p)^2} \qquad \ldots\ldots(2.13)$$

$$b = \frac{\sinh p + \sin p}{2p(\cosh p + \cos p)} + \frac{(1 + \cosh p \cos p)}{(\cosh p + \cos p)^2} \qquad \ldots\ldots(2.14)$$

and $p_1 = 2t/\delta$. Equation (2.11) agrees with the well-known expression for the eddy current loss, which is usually derived by a different method (see, for example, Russell 1904, p. 360). The variations of the coefficients a and b with p are shown in figure 1, and of the ratio a/b in figure 4.

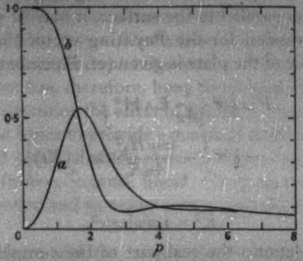


Figure 1. Reduced nuclear absorption coefficients as a function of reduced specimen thickness for a flat plate. The curves show the values of the coefficients a and b, which are such that the nuclear resonance power absorption per unit volume W_1 is given by $2W_1/\omega H_1^2 = a\chi' + b\chi''$. The plate thickness is $2t$; $p = 2t/\delta$, where δ is the skin depth (cf. equations (2.5), (3.1)). H_1 and ω are the amplitude and angular frequency of the radio-frequency field.

Before discussing the implications of these results it will be convenient to derive the corresponding results for a cylinder and a sphere.

(ii) *Cylinder.*

It may be shown that the magnetic field inside an infinitely long cylinder of radius t in a uniform alternating magnetic field, parallel to the axis, is given (cf. Smythe 1939, p. 392) by

$$H_\zeta(\rho) = \frac{H_1 \mathrm{I}_0(K\rho)}{\mathrm{I}_0(Kt)}. \qquad \ldots\ldots(2.15)$$

In this equation, cylindrical polar coordinates, ζ, ρ, ϕ, are used, I_0 is the modified Bessel function of the first kind and zero order, K is defined by (2.3), and, as is customary, the time factor has been omitted. From Maxwell's equations it may now be shown that the electric field inside the cylinder has only a ϕ-component, which is given by

$$E_\phi(\rho) = -\frac{cKH_1}{4\pi\sigma}\,\frac{\mathrm{I}_0'(K\rho)}{\mathrm{I}_0(Kt)}. \qquad \ldots\ldots(2.16)$$

Then, using the complex Poynting vector (cf. (2.2)), the power absorption per unit volume may be shown to be

$$W = \frac{c^2H_1^2}{16\pi^2 t\sigma}\mathcal{R}\left\{\frac{K\mathrm{I}_0'(Kt)}{\mathrm{I}_0(Kt)}\right\}. \qquad \ldots\ldots(2.17)$$

From the definition of K, and noting that $\alpha \ll 1$, $\beta \ll 1$,

$$\frac{K\mathrm{I}_0'(Kt)}{\mathrm{I}_0(Kt)} \simeq \mathrm{Y}_0 + \mathrm{Y}_1 \qquad \ldots\ldots(2.18)$$

where

$$\mathrm{Y}_0 = (z\mathrm{I}_0')/(t\mathrm{I}_0) \qquad \ldots\ldots(2.19)$$

$$Y_1 = Y_0 \left\{1 + \left(\frac{zI_0''}{I_0'}\right) - \left(\frac{zI_0'}{I_0}\right)\right\}(\alpha - i\beta) \qquad \ldots\ldots(2.20)$$

where $z = (2i)^{1/2}(t/\delta)$, and the argument of the Bessel functions is z. Introducing the Kelvin functions defined (Dwight 1947, p. 184) by

$$I_0(x\sqrt{i}) = \operatorname{ber} x + i \operatorname{bei} x \qquad \ldots\ldots(2.21)$$

it may be shown from equations (2.17) to (2.21) that W may be expressed in the form (2.1) with

$$\frac{8\pi W_0}{\omega H_1^2} = \frac{2}{s}\left\{\frac{\operatorname{ber} s \operatorname{ber}' s + \operatorname{bei} s \operatorname{bei}' s}{\operatorname{ber}^2 s + \operatorname{bei}^2 s}\right\} \qquad \ldots\ldots(2.22)$$

and W_1 in the form (2.12) with

$$a = \frac{(\operatorname{ber}^2 s - \operatorname{bei}^2 s)(\operatorname{bei}'^2 s - \operatorname{ber}'^2 s) - 4 \operatorname{ber} s \operatorname{ber}' s \operatorname{bei} s \operatorname{bei}' s}{(\operatorname{ber}^2 s + \operatorname{bei}^2 s)^2} \qquad \ldots\ldots(2.23)$$

$$b = 1 - \frac{2(\operatorname{ber} s \operatorname{ber}' s + \operatorname{bei} s \operatorname{bei}' s)(\operatorname{ber} s \operatorname{bei}' s - \operatorname{ber}' s \operatorname{bei} s)}{(\operatorname{ber}^2 s + \operatorname{bei}^2 s)^2} \qquad \ldots\ldots(2.24)$$

where $s = \sqrt{2}(t/\delta) = p/\sqrt{2}$. Equation (2.22) agrees with the corresponding expression given by Russell (1904, p. 374). Values of a and b, calculated from (2.23), (2.24) by means of the table of Kelvin functions given by Dwight (1947), are shown as functions of $p = s\sqrt{2}$ in figure 2, and the corresponding values of a/b in figure 4.

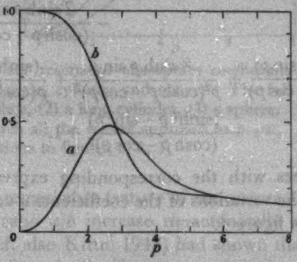


Figure 2. Reduced nuclear resonance absorption coefficients as a function of reduced specimen diameter for a long cylinder. The cylinder diameter is $2t$; $p = 2t/\delta$. For the definitions of the other symbols see figure 1.

(iii) *Sphere.*

Casimir's (1947) treatment of a conducting sphere in an alternating magnetic field leads to the following expression for the power absorption per unit volume in a sphere of radius t (cf. also Bloembergen 1952):

$$W = -\frac{3\omega H_1^2}{8\pi} \mathscr{I}\left\{\frac{\mu v(t) - 1}{\mu v(t) + 2}\right\} \qquad \ldots\ldots(2.25)$$

where

$$v(t) = \frac{2\{Kt - \tanh(Kt)\}}{(K^2t^2 + 1)\tanh(Kt) - Kt} \qquad \ldots\ldots(2.26)$$

and K and μ are defined in (2.3) and (2.6). In analysing (2.25), it is convenient to write

$$\left.\begin{aligned}\mu&=1+2(\alpha-i\beta)=1+\mu_1\\ v&=v_0+v_1\end{aligned}\right\}\qquad\qquad\ldots\ldots(2.27)$$

where v_0 is independent of α and β, $\mu_1 \ll 1$ and $v_1 \ll v_0$, since $\alpha \ll 1$, $\beta \ll 1$. Then, neglecting products of μ_1 and v_1

$$\frac{\mu v-1}{\mu v+2}\simeq\frac{v_0-1}{v_0+2}+\frac{3(v_1+v_0\mu_1)}{(v_0+2)^2}.\qquad\qquad\ldots\ldots(2.28)$$

Also from (2.26), (2.27)

$$v_0=\frac{2(z-\psi)}{(1+z^2)\psi-z}\qquad\qquad\ldots\ldots(2.29)$$

$$v_1=\frac{2z^2(z^2\psi^2+2\psi^2-z^2-z\psi)(\alpha-i\beta)}{\{(1+z^2)\psi-z\}^2}\qquad\qquad\ldots\ldots(2.30)$$

where z and ψ are defined in (2.10). Then from equations (2.25) to (2.30), it may be shown that W may be expressed in the form (2.1) with

$$\frac{8\pi W_0}{\omega H_1^2}=\frac{9}{p^2}\left\{\frac{p(\sinh p+\sin p)}{2(\cosh p-\cos p)}-1\right\}\qquad\qquad\ldots\ldots(2.31)$$

and W_1 in the form (2.12) with

$$\frac{4a}{9}=\frac{8(\sinh p-\sin p)}{p^3(\cosh p-\cos p)}-\frac{4(\sinh^2 p-\sin^2 p)}{p^2(\cosh p-\cos p)^2}+\frac{(\sinh p+\sin p)}{p(\cosh p-\cos p)}-\frac{2\sinh p\sin p}{(\cosh p-\cos p)^2}\qquad\qquad\ldots\ldots(2.32)$$

$$\frac{4b}{9}=\frac{8}{p^4}-\frac{8(\sinh p+\sin p)}{p^3(\cosh p-\cos p)}+\frac{8\sinh p\sin p}{p^2(\cosh p-\cos p)^2}+\frac{(\sinh p-\sin p)}{p(\cosh p-\cos p)}-\frac{(\sinh^2 p-\sin^2 p)}{(\cosh p-\cos p)^2}+1.\qquad\qquad\ldots\ldots(2.33)$$

Equation (2.31) agrees with the corresponding expression given by Smythe (1939, p. 414, 3). The variations of the coefficients a and b with p are shown in figure 3, and of a/b in figure 4.

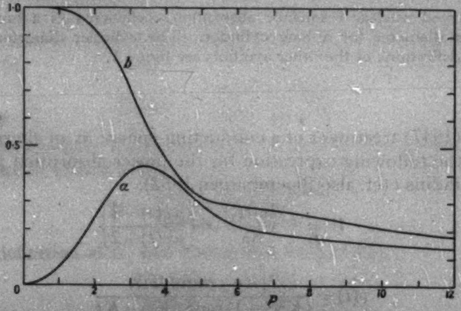


Figure 3. Reduced nuclear resonance absorption coefficients as a function of reduced specimen diameter for a sphere. The diameter of the sphere is $2t$; $p=2t/\delta$. For the definitions of the other symbols see figure 1.

(iv) ***Line shapes.***

In all three cases considered above the part of the power absorption depending on χ' and χ'', i.e., that part of the absorption which is measured in a nuclear resonance experiment, may be expressed as

$$\begin{aligned} W_1 &= \tfrac{1}{2}\omega H_1^2(a\chi' + b\chi'') \\ &= \tfrac{1}{2}\omega H_1^2 b\{\chi'' + (a/b)\chi'\}. \end{aligned} \qquad \ldots\ldots(2.34)$$

Consequently, the observable nuclear absorption will be proportional to a linear combination of χ'' and χ', and a measure of the admixture of the 'dispersive' component χ' is given by the ratio a/b. The dependence of this ratio on reduced specimen size for plate, cylinder and sphere is shown in figure 4.

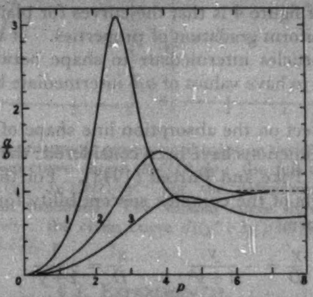


Figure 4. Ratio of nuclear resonance absorption coefficients as a function of reduced specimen size for plate, cylinder and sphere. The full curves show the values of a/b for (1) a flat plate, (2) a long cylinder, (3) a sphere. The broken line shows the limiting value which all the curves approach as $p \to \infty$. For the definitions of a, b and p see the captions to figures 1, 2, 3.

Perhaps the most remarkable feature of the curves shown in figure 4 is that in no case does the ratio a/b increase monotonically with p. Previous work (Bloembergen 1952, cf. also Kittel 1948) had shown that, for the plate and the sphere, $a/b \to 0$ as $p \to 0$, and $a/b \to 1$ as $p \to \infty$, but these earlier results gave no indication of the ranges over which these two limiting expressions would provide good approximations, nor was there any suggestion of a maximum in the curve of a/b as a function of p. Dyson (1955, Section VII), in the course of a theoretical treatment of electron spin resonance absorption in metals, has given expressions which, incidentally, should lead to values of a and b appropriate to nuclear resonance in a flat plate. It might be inferred from Dyson's discussion of these expressions that the approximation $a/b = 0$ (in the present notation) would then be reasonably accurate up to $p = 4$, in marked disagreement with the results of the present treatment. In fact, however, in the appropriate limit (i.e. in Dyson's notation, $R \to \infty$; corresponding to the absence of diffusion) Dyson's expressions may be reduced to the form (2.12), (2.13), (2.14) for the particular case in which χ', χ'' correspond to Lorentz shape functions (cf. (2.36) below).

As can be seen from figure 4, for the plate the increase of a/b above the limiting value of 1 is very pronounced. Indeed, in the region of $p = 2{\cdot}6$, a/b is greater than 3. This implies that for foils with the corresponding reduced thickness an absorption mode measurement would give a resonance curve determined largely by the 'dispersive' component of the susceptibility χ'. For the cylinder the curve of a/b again shows a maximum, but this is not so large as for the plate. The maximum for the sphere is even smaller, but an interesting feature here is that the function approaches the limiting value, for $p \to \infty$, very slowly, as may be seen from the following series expansion derived from (2.32), (2.33):

$$\frac{a}{b} = 1 - \frac{4}{p} + \frac{16}{p^2} - \frac{40}{p^3} + \frac{160}{p^4} + O\left(\frac{1}{p^5}\right). \qquad \ldots\ldots(2.35)$$

(The corresponding series expansions, in powers of p and of $1/p$, for W_0, a and b for the plate, cylinder and sphere, which may be derived from the explicit expressions given above, have also been found useful.)

A further feature of figure 4 is that the curves for plate, cylinder and sphere show a more or less uniform gradation of properties. It would seem reasonable therefore to expect particles intermediate in shape between spheres and long cylinders, for example, to have values of a/b intermediate between curves (2) and (3) in figure 4.

To illustrate the effect on the absorption line shape of introducing a term in χ' two idealized shape functions have been considered: the Lorentz and Gaussian functions discussed by Pake and Purcell (1948). For the present purpose the real and imaginary parts of the complex susceptibility for the Lorentz function may be written in the form

$$\frac{\chi'}{D} = -\frac{y}{1+y^2} \qquad \frac{\chi''}{D} = \frac{1}{1+y^2}. \qquad \ldots\ldots(2.36)$$

In these equations $y = 2\pi T_2(\nu - \nu_0) = T_2\,\gamma(H_0 - H)$, D and T_2 are constants, γ is the nuclear gyromagnetic ratio, ν and ν_0 are the frequency of the applied radio-frequency field and the resonant frequency, if resonance is measured at constant applied field, H and H_0 are the applied static magnetic field and the resonant magnetic field, if resonance is measured at a constant frequency. In figure 5 (*a*) curves are shown of

$$\frac{d}{dy}\left(\frac{\chi''}{D}\right), \quad \frac{d}{dy}\left(\frac{\chi''+\chi'}{D}\right) \quad \text{and} \quad \frac{d}{dy}\left(\frac{\chi''+3\chi'}{3D}\right)$$

as functions of y, calculated from (2.36). These correspond to absorption derivative curves (such as may be observed experimentally, cf. § 3) with $a/b = 0$, 1 and 3, respectively.

Similarly, for the Gaussian function

$$\frac{\chi'}{D} = -\frac{2}{\sqrt{\pi}}\exp(-\eta^2)\int_0^\eta \exp(x^2)\,dx \qquad \frac{\chi''}{D} = \exp(-\eta^2) \qquad \ldots\ldots(2.37)$$

where $\eta = 2\sqrt{\pi}\,T_2(\nu - \nu_0) = (T_2/\sqrt{\pi})\gamma(H_0 - H)$. The associated absorption derivatives

$$\frac{d}{d\eta}\left(\frac{\chi''}{D}\right), \quad \frac{d}{d\eta}\left(\frac{\chi''+\chi'}{D}\right) \quad \text{and} \quad \frac{d}{d\eta}\left(\frac{\chi''+3\chi'}{3D}\right)$$

are shown as functions of η in figure 5 (*b*).

As can be seen from figure 5, the derivative curves for $a/b = 0$ are antisymmetrical, but as a/b increases this property is lost, and the 'central' zeros move to positions corresponding to lower values of the frequency ν. Also, a second maximum occurs at a positive value of the variable. Although, in these respects, the properties of the Lorentz and Gaussian lines are qualitatively similar, there are obvious quantitative differences, but these may best be considered in relation to the experimentally observed curves (cf. §§ 3, 4).

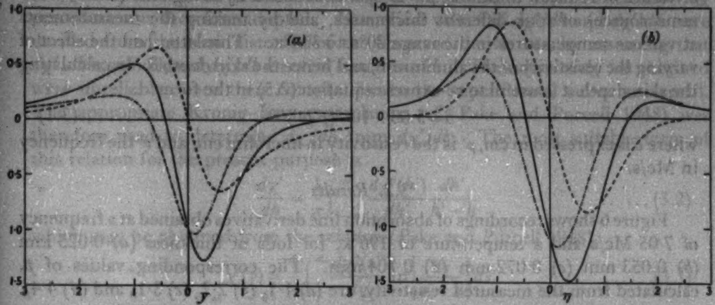


Figure 5. Theoretical nuclear magnetic resonance absorption derivative curves. (*a*) Lorentz function. The broken curve shows $d(\chi''/D)/dy$; the full curve $d\{(\chi''+\chi')/D\}/dy$; the chain curve $d\{(\chi''+3\chi')/3D\}/dy$. For the definitions of χ', χ'' and y see (2.36). (*b*) Gaussian function. The broken curve shows $d(\chi''/D)/d\eta$; the full curve $d\{(\chi''+\chi')/D\}/d\eta$; the chain curve $d\{(\chi''+3\chi')/3D\}/d\eta$. For the definitions of χ', χ'' and η see (2.37).

§ 3. EXPERIMENTAL

3.1. *Method and Apparatus*

In order to test the applicability of the theoretical treatment of § 2 an experimental study has been made of the ^{27}Al resonance in foils of pure aluminium. Foils were chosen as being of particular interest in view of the fact that the theoretical results for this shape of specimen deviate most strikingly from those previously expected (cf. figure 4).

The apparatus consisted of a marginal oscillator of the type described by Pound and Knight (1950). The magnetic field was modulated at 200 c/s with an amplitude less than the line width, and resonances were traversed by slowly and continuously changing the field. Since only amplitude modulation of the radio-frequency oscillation was detected, the apparatus was sensitive only to the absorption mode. Derivatives of absorption lines, with respect to magnetic field, were traced out by a recording milliammeter which responded to the output of a phase-sensitive detector. The magnetic field was measured by observation of the frequency of a proton nuclear magnetic resonance, by means of a second marginal oscillator which could be rapidly and easily substituted for the first. The apparatus also included a 'calibrator' circuit (Pound 1952), which could be used to produce controllable amounts of absorption of the radio-frequency power. By comparison of the calibrator signals with the nuclear signals, relative magnitudes of the latter could be measured irrespective of any changes in the gain of the amplifiers, or in the sensitivity of the oscillating circuit.

Samples were made by interleaving about fifty sheets of annealed aluminium foil (purity 99·99%), 2 cm long, 0·9 cm wide, and of various thicknesses (see below), with sheets of polythene, 0·05 mm thick. The samples were placed centrally in a radio-frequency coil of length 3 cm, so that the radio-frequency magnetic field was closely parallel to the planes of the sheets throughout their lengths, and conditions approximated to those assumed in §2 (i). The required range of values of reduced thickness, $p=2t/\delta$, was obtained by using foils (all from the same source) of four different thicknesses, and by making the measurements at various temperatures in the range 20°K to 300°K. This latter had the effect of varying the resistivity of the aluminium, and hence the skin depth δ. In calculating the skin depth it is useful to re-express equation (2.5) in the form

$$\delta = 5{\cdot}03 \times 10^{-2}(\rho'/\nu')^{1/2} \qquad \ldots\ldots(3.1)$$

where δ is expressed in cm, ρ' is the resistivity in microhm cm, and ν' the frequency in Mc/s.

3.2. *Results*

Figure 6 shows recordings of absorption line derivatives obtained at a frequency of 7·05 Mc/s and a temperature of 196°K, for foils of thickness (*a*) 0·025 mm (*b*) 0·053 mm (*c*) 0·072 mm (*d*) 0·104 mm. The corresponding values of p, calculated from the measured resistivity, are (*a*) $1{\cdot}1_0$ (*b*) $2{\cdot}2_8$ (*c*) $3{\cdot}1_0$ and (*d*) $4{\cdot}4_8$.

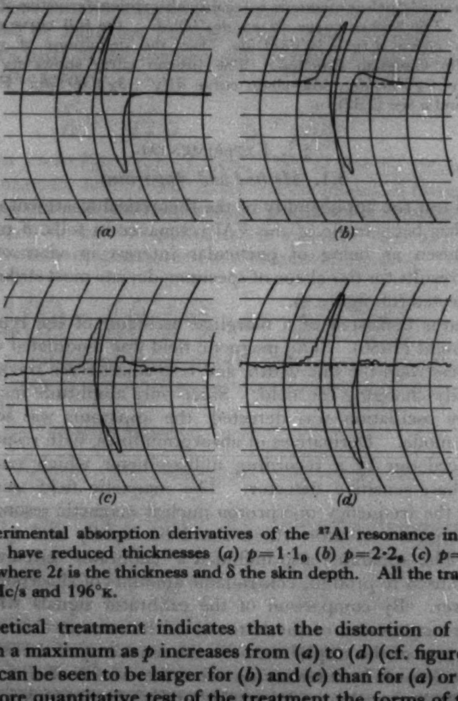


Figure 6. Experimental absorption derivatives of the ^{27}Al resonance in aluminium foils. The foils have reduced thicknesses (*a*) $p=1{\cdot}1_0$ (*b*) $p=2{\cdot}2_8$ (*c*) $p=3{\cdot}1_0$ (*d*) $p=4{\cdot}4_8$. $p=2t/\delta$, where $2t$ is the thickness and δ the skin depth. All the traces were obtained at 7·05 Mc/s and 196°K.

The theoretical treatment indicates that the distortion of the derivatives passes through a maximum as p increases from (*a*) to (*d*) (cf. figure 4), and indeed the distortion can be seen to be larger for (*b*) and (*c*) than for (*a*) or (*d*).

For any more quantitative test of the treatment the forms of the true nuclear absorption and dispersion derivatives, $d\chi''/dh$ and $d\chi'/dh$ (where $h=H_0-H$, and H_0 corresponds to the centre of the true absorption line), are required. By inspection, it was clear that the observed derivatives could not be represented at all closely by the Lorentz or Gaussian functions. Nor could either $d\chi''/dh$ or $d\chi'/dh$ be observed directly with any of the available foils. The following procedure was therefore adopted. Numerical values of $d\chi''/dh$ and $d\chi'/dh$ were derived from the least distorted of the observed derivatives, by assuming that $d\chi''/dh$ is an odd, and $d\chi'/dh$ an even, function of h. The values of $d\chi'/dh$ obtained were unreliable, since their contribution to the experimental curve was small. The appropriate Kronig–Kramers relation (cf. Pake and Purcell 1948) was therefore used to determine $d\chi'/dh$ from $d\chi''/dh$. The most suitable form of this relation for the present purpose is

$$\frac{d\chi'}{dh} = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{d\chi''(h')}{dh'}\,\frac{dh'}{h'-h} \qquad \ldots\ldots(3.2)$$

which may be obtained from the relation (Pake and Purcell 1948)

$$\chi'(\nu) = \frac{1}{\pi}\int_{0}^{\infty} \frac{d\chi''(\nu')}{d\nu'} \ln\left|\frac{\nu_0}{\nu'-\nu}\right| d\nu' \qquad \ldots\ldots(3.3)$$

by introducing the variable h, defined by $2\pi(\nu-\nu_0)=\gamma(H_0-H)=\gamma h$, and differentiating with respect to h. The lower limit of integration in (3.2) has been written as $-\infty$ instead of $-2\pi\nu_0/\gamma$, but this introduces no appreciable error.

The results obtained for $d\chi''/dh$ and $d\chi'/dh$ from this analysis depended on the value assumed for H_0. The function $\{a\,d\chi'/dh + b\,d\chi''/dh\}$ was calculated, therefore, for a range of values of H_0, using the values of a and b, from (2.13), (2.14), corresponding to the appropriate value of p. The true values of H_0, $d\chi''/dh$ and $d\chi'/dh$ were then taken to be those for which the calculated function agreed most closely with the initial experimental trace. The agreement between the final calculated function and the experimental trace was comparable with that between different experimental traces obtained under the same conditions.

The absorption derivative corresponding to any value of p could now be obtained by using the calculated functions $d\chi''/dh$ and $d\chi'/dh$ and the appropriate pair of values of a and b. Quantitative comparisons between some of the characteristic features of derivatives obtained in this way and those of the corresponding experimentally observed derivatives are shown in figures 7, 8, 9. It must be emphasized that the curves in these figures apply only to systems characterized by the particular derivative functions $d\chi''/dh$ and $d\chi'/dh$ obtained above, although qualitatively similar curves would be expected for other pairs of functions.

The curves in figure 7 show the calculated ratios of extreme positive values B, C, of the derivative, to extreme negative value A, as functions of the reduced specimen thickness p. These ratios constitute the most readily obtainable measure of line shape. The points show experimental results for foils of four different thicknesses, at a series of temperatures. It can be seen that the trend of these results follows that of the curves; in particular, the maximum deviation from the values $B/A=1$, $C/A=0$ occurs near $p=2{\cdot}6$ as predicted. There is, however, some discrepancy, which is probably attributable, in part at least, to the inhomogeneity of the steady magnetic field over the sample, amounting to about 1·5 gauss. A further discrepancy could arise through any change in $d\chi''/dh$ and $d\chi'/dh$ with temperature, other than a change by a factor independent of h. It should also be mentioned that the agreement might have been improved by the choice of a different experimental derivative as the starting point of the analysis.

The curve in figure 8 shows the ratio of the calculated shift h_p, of the apparent centre of the line, to the half-width h_w. The value of the half-width, defined as half the separation of the extreme values of the undistorted absorption derivative, was 4·4 gauss. The curve passes through a maximum near $p = 2{\cdot}6$ corresponding to the maximum in a/b. The points represent experimental values of h_p/h_w for foil of thickness 0·025 mm at temperatures of 300°K, 77°K and 20°K. A small change in the position of the true absorption line in this temperature range has been taken into account. Measurements made on a sample of commercial aluminium powder, which gave symmetrical resonances, showed that this shift of the true resonance amounted to a decrease in the Knight (1949) shift by $(4 \pm 2)\%$ of itself as the temperature was reduced from 300°K to 20°K. This result is in qualitative agreement with that of Knight (1954) who does not, however, give a numerical value. The predicted maximum in the shift is confirmed although

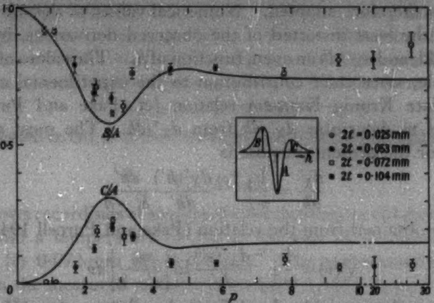


Figure 7. Ratios of peak values of the derivative of the ^{27}Al resonance in aluminium foils as a function of reduced thickness. *A*, *B* and *C* are defined in the inset figure. $p = 2t/\delta$, where $2t$ is the foil thickness and δ is the skin depth.

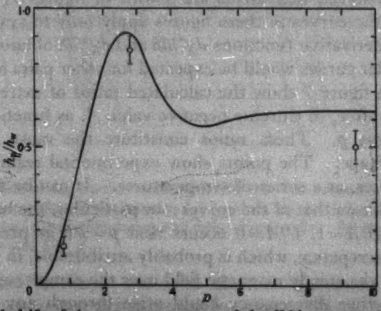


Figure 8. Reduced shift of the apparent centre of the ^{27}Al resonance in aluminium foil as a function of reduced thickness. h_p is the shift and h_w the half-width of the absorption line; $p = 2t/\delta$, where the thickness $2t$ is 0·025 mm, and δ is the skin depth.

all the three experimental points fall below the theoretical values. The reasons for the discrepancy are probably the same as those mentioned in connection with figure 7.

The theoretical expression for the absorption derivative may be written in the form (cf. (2.34))

$$b\left\{\frac{d\chi''}{dh}+\left(\frac{a}{b}\right)\frac{d\chi'}{dh}\right\}.$$

The characteristics of the derivative considered so far are independent of the variation of the coefficient b which appears outside the brackets in this expression, so that they involve the coefficients a and b only as a ratio. A test of the whole expression is provided by consideration of the variation of the amplitude of the derivative with reduced foil thickness p. A complication arises however if p is varied by changing the temperature, since χ' and χ'' are changed also, being inversely proportional to the absolute temperature. This must be taken into account in a comparison of calculated and experimental derivative amplitudes.

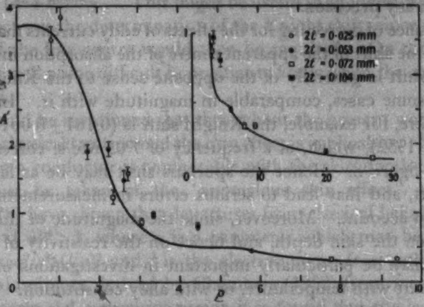


Figure 9. Reduced extreme negative value of the derivative of the ^{27}Al resonance in aluminium foils as a function of reduced foil thickness. $A'=AT/300$, where T is the absolute temperature. For the definitions of the other symbols see figure 7.

The curve in figure 9 shows the variation with p of the extreme negative value A' of the calculated absorption derivative. The curve refers to a fixed volume of metal, and $d\chi'/dh$, $d\chi''/dh$ have been assumed to retain the (room-temperature) values obtained as described above. The points represent experimental values multiplied by the factor $T/300$, where T is the absolute temperature of the specimen. This factor takes account of the temperature variation of $d\chi'/dh$ and $d\chi''/dh$. There is a rather large experimental scatter, but the predicted variation of A' with p is essentially confirmed.

§ 4. DISCUSSION

The degree of agreement between the theoretical and experimental results shown in § 3 is sufficient to indicate the general applicability of the results given in § 2. These results may now be used for foils, wires and spherical particles to give precise criteria for deciding, in any particular case, whether the effect of eddy currents will be appreciable.

Where eddy current effects are not negligible, corrections must be applied to the observed resonance curves in order to determine, for example, the true line centre and intensity. If the form of $\chi''(h)$ is known, then that of $\chi'(h)$ may be derived by means of the appropriate Kronig–Kramers relation, or vice versa, and the required correction to the observed curve may be obtained in a straightforward manner from the results in § 2. If however the form of $\chi''(h)$ is not known, but it may be assumed to be an even function of h, and $\chi'(h)$ consequently an odd function, the procedure outlined in § 3 may be adopted. When the form of $\chi''(h)$ is not known, and it may not be assumed to be an even function of h, recourse must be made to a process of trial. Functions $\chi''(h)$ and $\chi'(h)$ must be found which satisfy the Kronig–Kramers relations, and are such that when added in the proportions appropriate to the specimen used (§ 2) they reproduce the observed curve. In every case, in order to apply the results of § 2, the electrical conductivity and the appropriate dimensions of the specimen must be known. From this point of view, the use of foils or wires is to be preferred to that of powders since usually neither the shapes nor sizes of the constituent particles of a powder are known with any precision.

The importance of correcting for the effects of eddy currents may be illustrated by considering the shift of the apparent centre of the absorption line (cf. figure 8). This spurious shift is generally of the opposite sense to the Knight (1949) shift and is also, in some cases, comparable in magnitude with it. In aluminium at room temperature, for example, the Knight shift is $(0{\cdot}161 \pm 0{\cdot}001)\%$ (Gutowsky and McGarvey 1952), which, at a frequency of 7·05 Mc/s, corresponds to 2·44 in the units of figure 8. Hence the spurious shift may be as large as 40% of the Knight shift, and may lead to serious errors in measurements of the latter if not taken into account. Moreover, since the magnitude of the eddy current effect depends on the skin depth, and hence on the resistivity of the specimen, the correction may be particularly important in investigations of the variation of the Knight shift with temperature, or with alloy constitution.

In connection with the experimental results considered in § 3, it is interesting to note that the absorption derivative associated with $\chi''+\chi'$, for example, is not so obviously different from that associated with χ'' as might have been expected from consideration of the Lorentz and Gaussian functions (cf. figure 5). A measure of the distortion, due to eddy currents, of an otherwise antisymmetrical absorption derivative is given by the deviation of the ratio B/A (see figure 7) from unity. For the Lorentz and Gaussian derivatives of $\chi''+\chi'$ the values of B/A are 0·39 and 0·55 respectively, while for the line shape found experimentally in the present work the value is about 0·75, i.e. much closer to unity. Furthermore, the distortion of the absorption line itself may be expected to be even less obvious than that of the derivative, and yet the shift of the line centre and the change in intensity may still be significant (cf. figures 8, 9). For these reasons it would obviously be difficult to draw conclusions about the extent of the effect of eddy currents merely from the appearance of an observed absorption curve. Such considerations as these are probably relevant to the results of Redfield (1956), who found that the absorption curves for copper and aluminium powders were 'almost symmetrical' although the particle size was much greater than the skin depth. It may also be noted in this connection that, for spherical particles, a/b approaches unity very slowly with increasing particle size, and this again would tend to reduce the distortion of the absorption curve.

The present paper has been concerned largely with the theoretical and experimental aspects of the observation of nuclear resonance by absorption measurements. The results of the treatment may however be readily extended to apply to other methods of measurement. In § 2 expressions have been derived for the power absorption in a specimen in an alternating field, but these expressions may also be regarded as giving a measure of the component of magnetic flux in quadrature with the applied field (cf. the discussion of Andrew 1955, § 2.8). That is, the component of the induced magnetization in quadrature with the applied field is proportional to $a\chi' + b\chi''$. It may then be shown that the corresponding in-phase component of induced magnetization is proportional to $b\chi' - a\chi''$. If, therefore, a bridge circuit or a nuclear induction apparatus (cf. Andrew 1955, Chapter 3) is adjusted to detect the absorption mode the nuclear signal will be proportional to $a\chi' + b\chi''$; if adjusted to detect the dispersion mode the nuclear signal will be proportional to $b\chi' - a\chi''$. It may also be mentioned that it is possible, in principle, by combining the absorption and dispersion modes, for instance by means of a bridge circuit, to obtain a signal proportional either to $(a^2+b^2)^{1/2}\chi''$ or to $(a^2+b^2)^{1/2}\chi'$. This might be useful in an investigation in which the principal object was the determination of the Knight shift, since no correction to the position of the line centre would then be required. On the other hand, however, it might well prove difficult in practice to know when the correct balance had been attained.

Finally, it should be mentioned that throughout the present work it has been assumed that the classical treatment of the skin effect is applicable. If the electron mean free path in a conductor is comparable with the skin depth this is not so, and account must be taken of the 'anomalous skin effect'. Reuter and Sondheimer (1948) have shown that departures from classical behaviour are then appreciable only if $\alpha^{1/6} > 1$, where α is a parameter which may be expressed in the form $\alpha = (3\pi\omega\sigma^2/c^2)(l/\sigma)^2$ and l is the electron mean free path. Using the value of l/σ for aluminium derived by Chambers (1952) from measurements of surface impedance in the microwave region it is found that the largest value of $\alpha^{1/6}$ arising in the present work is about 1·0 at 20°K. Consequently, no significant departures from the results of the classical treatment are expected, and none are in fact observed. In the future, however, it is hoped to extend some of the measurements to liquid helium temperatures, and it might then be necessary to modify the theoretical treatment to take into account the anomalous skin effect.

ACKNOWLEDGMENTS

We are indebted to Professor E. C. Stoner for his interest and encouragement. We also wish to thank Dr. F. E. Hoare and his collaborators for the provision of low temperature facilities. One of us (A. C. C.) is grateful to the Department of Scientific and Industrial Research for the award of a maintenance allowance.

REFERENCES

ANDREW, E. R., 1955, *Nuclear Magnetic Resonance* (Cambridge: University Press).
BLOEMBERGEN, N., 1952, *J. Appl. Phys.*, **23**, 1383.
CASIMIR, H. B. G., 1947, *Philips Res. Rep.*, **2**, 42.
CHAMBERS, R. G., 1952, *Proc. Roy. Soc.* A, **215**, 481.
DWIGHT, H. B., 1947, *Tables of Integrals and other Mathematical Data*, 2nd edn. (New York: Macmillan).

DYSON, F. J., 1955, *Phys. Rev.*, **98**, 349.

GUTOWSKY, H. S. and MCGARVEY, B. R., 1952, *J. Chem. Phys.*, **20**, 1472.

KITTEL, C., 1948, *Phys. Rev.*, **73**, 155.

KNIGHT, W. D., 1949, *Phys. Rev.*, **76**, 1259 ; 1954, *Ibid.*, **96**, 861.

PAKE, G. E. and PURCELL, E. M., 1948, *Phys. Rev.*, **74**, 1184.

POUND, R. V., 1952, *Progr. Nucl. Phys.*, **2**, 21.

POUND, R. V. and KNIGHT, W. D., 1950, *Rev. Sci. Instrum.*, **21**, 219.

REDFIELD, A. G., 1956, *Phys. Rev.*, **101**, 67.

REUTER, G. E. H., and SONDHEIMER, E. H., 1948, *Proc. Roy. Soc.* A, **195**, 336.

RUSSELL, A., 1904, *A Treatise on the Theory of Alternating Currents*, I (Cambridge : University Press).

SMYTHE, W. R., 1939, *Static and Dynamic Electricity* (New York : McGraw-Hill).

The Influence of Condenser Aperture on the Resolution of a Pair of Parallel Lines

By A. M. GOODBODY
Technical Optics Section, Imperial College, S.W.7

Communicated by H. H. Hopkins; MS. received 22nd June 1956, and in final form 17th January 1957

Abstract. An expression is found for the distribution of light in the image of two parallel lines which are illuminated by a condenser of numerical aperture equal to σ times that of the imaging objective. The intensity distributions are plotted for various values of σ and of the separation of the lines. The resolving power of the system is written $\xi = K\lambda/\text{N.A.}$ and K is plotted as a function of σ.

§ 1. Introduction

The physical limits of resolution for two parallel transmitting slits have been found by Osterberg (1950) for the case in which the illumination is provided by light issuing from a condenser diaphragm whose opening is a slit orientated at right angles to the object slits. This type of illumination he has called lateral slit illumination. Further, the physical limits of resolution for pairs of like and unlike particles have been found for the case of a circular condenser.

In the above work, the criterion of resolution employed was one due to Sparrow (1916), namely that on the limit of resolution $\partial^2 I/\partial x^2 = 0$ at a point mid-way between the two geometric images, I being the intensity and x being a coordinate measured along the line joining the images. In the work that follows, however, the Rayleigh criterion of resolution has been adopted. This requires that the minimum of intensity midway between the geometrical images is 20% smaller than the intensity at the maxima of the summed intensities.

The time averaged intensity in the image of two objects illuminated by incoherent light is found by adding the intensities due to each object separately. In the case of coherent light it is necessary to introduce a product term depending on the phase difference between the disturbances from the two objects. In general, however, the light will be partially coherent and the product term must be modified by a factor depending on the degree of correlation between the disturbances from the two objects. This factor has been termed the phase-coherence factor (Hopkins 1951). Thus if two disturbances of complex amplitudes $A_1 \exp(i\psi_1)$, $A_2 \exp(i\psi_2)$ are superposed at a given point the resulting time-averaged intensity will be given by

$$I = A_1^2 + A_2^2 + 2A_1A_2\mathscr{R}[\Gamma_{12} \exp\{i(\psi_1 - \psi_2)\}]$$

where Γ_{12} is the phase-coherence factor,† and $\mathscr{R}$ denotes, as usual, the real part.

---

† This expression is strictly valid only for perfectly monochromatic light. For light sources with a finite range of frequencies the right hand side has to be integrated with respect to frequency, but the result is unchanged for narrow frequency bands when only small path differences are involved.

Hopkins (1953) has shown that coherence and incoherence may conveniently be regarded as the special cases when $\sigma \to 0$ and $\sigma \to \infty$, where σ is the radius of an 'effective source'. With suitable units σ is merely equal to the ratio of the numerical aperture of the condenser to that of the objective. For two point disturbances the expression for the intensity in the image assumes a simple form (Hopkins and Barham 1950). The formula for the intensity distribution of two lines, however, requires an integration in which the effective source occurs as a factor in the integrand. Hopkins (1953) has calculated the influence of condenser aperture on the image of a single line and we now propose to extend this to the case of a pair of lines.

Following Hopkins and Barham we shall write $\xi = K\lambda/n \sin\alpha$ for the just resolvable separation of the two lines. K is a numerical factor which depends on σ, the ratio of the numerical aperture of the condenser to that of the objective. This latter is equal to $n \sin\alpha$, and λ is the vacuum wavelength of the light.

§ 2. Expression for the Intensity in the Image

If the incident and refracted parts of an aperture ray make angles α, α' with the axis and (ξ, η) are geometrical distances in the object, reduced coordinates for the object plane are defined by the expressions

$$u = \frac{2\pi}{\lambda}(n \sin\alpha)\xi, \qquad v = \frac{2\pi}{\lambda}(n \sin\alpha)\eta$$

where n is the refractive index in the object space. Similarly for the image plane

$$u' = \frac{2\pi}{\lambda}(n' \sin\alpha')\xi', \qquad v' = \frac{2\pi}{\lambda}(n' \sin\alpha')\eta'$$

from which it will be seen that $u' = u$, $v' = v$ if (u', v') is the geometrical image of (u, v).

If the object consists of two parallel transparent lines of negligible width, situated respectively at distances $+u_0$, $-u_0$ from the v axis, its structure may be specified by the expression

$$A(u) = \sqrt{(2\pi)}\{\delta(u-u_0) + \delta(u+u_0)\} \qquad \ldots\ldots(1)$$

where δ is a Dirac delta function. The Fourier transform of $A(u)$ is given by

$$a(s) = 2\cos u_0 s. \qquad \ldots\ldots(2)$$

Using the results obtained by Hopkins (1953) we may express the intensity in the image at the point (u', v') by means of the integral

$$B'(u', v') = 2\pi \iint_{-\infty}^{+\infty} \gamma(x, y)|\phi(x, y)|^2\, dx\, dy \qquad \ldots\ldots(3)$$

where

$$\phi(x, y) = \frac{1}{\sqrt{(2\pi)}} \exp(iyv') \int_{-\infty}^{+\infty} a(s-x) f(s, y) \exp(iu's)\, ds \qquad \ldots\ldots(4)$$

and $f(x, y)$ is the pupil function of the image forming system. This function is given by

$$\left.\begin{aligned} f(x, y) &= 1, \quad x^2+y^2 \leqslant 1 \\ &= 0, \quad x^2+y^2 > 1 \end{aligned}\right\} \qquad \ldots\ldots(5)$$

a circular aperture of radius $\sqrt{(x^2+y^2)} = 1$ being assumed. $\gamma(x, y)$ is the distribution of intensity in the effective source defined by Hopkins (1953).

Substitution of the expression for $a(s)$ in (4) leads to

$$\left.\begin{aligned} |\phi(x,y)|^2 = {} & \frac{2}{\pi}\left[\frac{\sin^2\{(u'-u_0)\sqrt{(1-y^2)}\}}{(u'-u_0)^2} + \frac{\sin^2\{(u'+u_0)\sqrt{(1-y^2)}\}}{(u'+u_0)^2}\right] \\ & + \frac{4}{\pi}\,\frac{\cos 2u_0x \sin\{(u'-u_0)\sqrt{(1-y^2)}\}\sin\{(u'+u_0)\sqrt{(1-y^2)}\}}{u'^2-u_0^2}, \quad |y| \leqslant 1 \\ = {} & 0, \qquad\qquad |y| > 1 \end{aligned}\right\} \quad \ldots\ldots(6)$$

since $f(x,y)=0$ for $y>1$. If the object is illuminated by a condenser system whose aperture is uniformlyi lluminated and whose numerical aperture is σ times that of the image forming system, the effective source is of uniform intensity inside a circle of radius σ, and we may let this intensity be equal to unity.

The first two terms in (6) are independent of x, and x occurs in the last term only as the factor $\cos 2u_0x$. The integration with respect to x in (3) may therefore be easily effected and leads to the factors

$$\left.\begin{aligned} & \int_{-\infty}^{+\infty} \gamma(x,y)\,dx = 2\int_0^{\sqrt{(\sigma^2-y^2)}} dx = 2\sqrt{(\sigma^2-y^2)} \\ & \text{and} \\ & \int_{-\infty}^{+\infty} \gamma(x,y)\cos 2u_0x\,dx = 2\int_0^{\sqrt{(\sigma^2-y^2)}} \cos 2u_0x\,dx = \frac{1}{u_0}\sin 2u_0\sqrt{(\sigma^2-y^2)} \end{aligned}\right\} \quad \ldots\ldots(7)$$

respectively.

Since B' is independent of v' but depends on u', u_0 and σ, we shall for convenience write $B'(u', u_0, \sigma)$ in place of $B'(u', v')$. From the relations (3), (6) and (7) we then obtain

$$\begin{aligned} B'(u', u_0, \sigma) = {} & 16\int_0^{\sigma} \sqrt{(\sigma^2-y^2)} \\ & \times\left[\frac{\sin^2\{(u'-u_0)\sqrt{(1-y^2)}\}}{(u'-u_0)^2} + \frac{\sin^2\{(u'+u_0)\sqrt{(1-y^2)}\}}{(u'+u_0)^2}\right] dy \\ & + 8\int_0^{\sigma}\left[\frac{\sin\{2u_0\sqrt{(\sigma^2-y^2)}\}[\cos\{2u_0\sqrt{(1-y^2)}\} - \cos\{2u'\sqrt{(1-y^2)}\}]}{u_0(u'-u_0)^2}\right] dy. \end{aligned} \quad \ldots\ldots(8)$$

The integration of the first term in (8) has been given by Hopkins, and leads to the result

$$\begin{aligned} 4\pi\sigma^2 \Bigg[& \left\{\frac{\sin^2(u'-u_0)}{(u'-u_0)^2} + \frac{\sin^2(u'+u_0)}{(u'+u_0)^2}\right\} \\ & + \sum_{n=2,4,\ldots}^{\infty} (-)^{\frac{1}{2}(n-2)} \left\{\frac{\mathrm{J}_n(2u'+2u_0)}{(u'+u_0)^2} + \frac{\mathrm{J}_n(2u'-2u_0)}{(u'-u_0)^2}\right\} \\ & \times \sum_{p=1}^{\frac{1}{2}n} (-)^p \frac{n^2(n^2-2^2)\ldots\ldots\{n^2-(2p-2)^2\}}{p!\,(p+1)!}\left(\frac{\sigma}{2}\right)^{2p}\Bigg]. \end{aligned} \quad \ldots\ldots(9)$$

To evaluate the first term of the second integral in (8) namely

$$I = \int_0^{\sigma} \sin\{2u_0\sqrt{(\sigma^2-y^2)}\}\cos\{2u'\sqrt{(1-y^2)}\}\,dy$$

we shall put $y=\sin\theta$, $\sigma=\sin\beta$, and the integral then becomes

$$I = \int_0^{\sigma} \sin\{2u_0\sqrt{(\sigma^2-\sin^2\theta)}\}\cos(2u'\cos\theta)\cos\theta\,d\theta. \quad \ldots\ldots(10)$$

The term $\cos(2u'\cos\theta)$ may be expended in a Jacobi series, and the cosines of multiples of θ which occur as coefficients are expanded in powers of $\sin\theta$ giving

$$I=\int_0^{\beta}\sin\{2u_0\sqrt{(\sigma^2-\sin^2\theta)}\}\Bigg\{\mathrm{J}_0(2u')+2\sum_{n=2,4,\ldots}^{\infty}(-)^{\frac{1}{2}n}\mathrm{J}_n(2u')$$
$$+2\sum_{n=2,4,\ldots}^{\infty}(-)^{\frac{1}{2}n}\mathrm{J}_n(2u')\sum_{p=1}^{\frac{1}{2}n}(-)^p\frac{n^2(n^2-2^2)\ldots.\{n^2-(2p-2)^2\}\sin^{2p}\theta}{(2p)!}\Bigg\}\cos\theta\,d\theta. \qquad \ldots\ldots(11)$$

Replacing $\sin\theta$ by $\sigma\sin\psi$ and expanding $\sin(2u_0\sigma\cos\psi)$ in a Jacobi series we obtain

$$I=2\int_0^{\pi/2}\sum_{m=0}^{\infty}(-)^m\mathrm{J}_{2m+1}(2u_0\sigma)\cos(2m+1)\psi\cos\psi$$
$$\times\Bigg\{\cos 2u'+2\sum_{n=2,4,\ldots}^{\infty}(-)^{\frac{1}{2}n}\mathrm{J}_n(2u')$$
$$\times\sum_{p=1}^{\frac{1}{2}n}(-)^p\frac{n^2(n^2-2^2)\ldots\{n^2-(2p-2)^2\}\sigma^{2p}\sin^{2p}\psi}{(2p)!}\Bigg\}\sigma\cos\psi\,d\psi. \qquad \ldots\ldots(12)$$

Writing $2\cos(2m+1)\psi\cos\psi$ as $\{\cos(2m+2)\psi+\cos 2m\psi\}$ and expanding these in powers of $\sin\theta$ and $\cos\theta$ we are left with integrals which can be written simply in terms of gamma functions.

Combining this result with expression (9) we finally obtain in place of (8)

$$B'(u',u_0,\sigma)=\frac{\sin^2(u'-u_0)}{(u'-u_0)^2}+\frac{\sin^2(u'+u_0)}{(u'+u_0)^2}+\frac{2\mathrm{J}_1(2u_0\sigma)}{u_0\sigma}\frac{\sin(u'-u_0)\sin(u'+u_0)}{(u'^2-u_0^2)}$$
$$+\sum_{n=2,4,\ldots}^{\infty}(-)^{\frac{1}{2}(n-2)}\sum_{p=1}^{\frac{1}{2}n}(-)^p\frac{n^2(n^2-2^2)\ldots\{n^2-(2p-2)^2\}}{p!(p+1)!}\left(\frac{\sigma}{2}\right)^{2p}$$
$$\times\left[\left\{\frac{\mathrm{J}_n(2u'+2u_0)}{(u'+u_0)^2}+\frac{\mathrm{J}_n(2u'-2u_0)}{(u'-u_0)^2}\right\}+2\frac{\mathrm{J}_n(2u')-\mathrm{J}_n(2u_0)}{u'^2-u_0^2}\right.$$
$$\times\sum_{m=0}^{\infty}(-)^m\frac{\mathrm{J}_{2m+1}(2u_0\sigma)}{u_0\sigma}\left\{\sum_{q=0}^{m}\frac{(-)^{p+q}(2m)!(\frac{1}{2})^{2m}}{(m-q)!(m+p)!(p+q)!}\binom{2p+2q}{2p}\right.$$
$$\left.\left.+\sum_{r=0}^{m+1}\frac{(-)^{p+r}(2m+2)!(\frac{1}{2})^{2m+2}}{(m-r+1)!(m+p+1)!(p+r)!}\binom{2p+2r}{2p}\right\}\right] \qquad \ldots\ldots(13)$$

from which the factor $4\pi\sigma^2$ has been omitted. (13) expresses the intensity in the image as a function of the condenser aperture and the separation of the two lines. We shall now show that (13) reduces to known results in special cases. Thus putting $u_0=0$ we find

$$B'(u',0,\sigma)=4\Bigg\{\left(\frac{\sin u'}{u'}\right)^2+\sum_{n=2,4,\ldots}^{\infty}(-)^{\frac{1}{2}(n-2)}\frac{\mathrm{J}_n(2u')}{u'^2}$$
$$\times\sum_{p=1}^{\frac{1}{2}n}(-)^p\frac{n^2(n^2-2^2)\ldots\{n^2-(2p-2)^2\}}{p!(p+1)!}\left(\frac{\sigma}{2}\right)^{2p}\Bigg\} \qquad \ldots\ldots(14)$$

which is the known result for a single line source. In the limit as $\sigma\to 0$, (13) simplifies to the form

$$B'(u',u_0,0)=\left\{\frac{\sin(u'-u_0)}{u'-u_0}+\frac{\sin(u'+u_0)}{u'+u_0}\right\}^2 \qquad \ldots\ldots(15)$$

which is the correct result for the case of coherent illumination.

When $\sigma=1$, equation (8) is easily integrated and leads to the result

$$B'(u',u_0,1)=2\pi\Bigg\{\frac{1-\mathrm{J}_0(2u'-2u_0)+\mathrm{J}_2(2u'-2u_0)}{(u'-u_0)^2}+\frac{1-\mathrm{J}_0(2u'+2u_0)+\mathrm{J}_2(2u'+2u_0)}{(u'+u_0)^2}$$
$$+\frac{\mathrm{J}_1(4u_0)+\mathrm{J}_1(2u'-2u_0)-\mathrm{J}_1(2u'+2u_0)}{u_0(u'^2-u_0^2)}\Bigg\}. \qquad \ldots\ldots(16)$$

If u_0 is now allowed to tend to zero, (16) simplifies to give

$$B'(u_0, 0, 1) = \frac{16\pi}{u'^2}\{1 - J_0(2u') + J_2(2u')\} \qquad \ldots\ldots(17)$$

which is again the known result for a single line source when $\sigma = 1$ (Hopkins 1953).

If we let σ tend to infinity the illumination is effectively incoherent and $\gamma(x, y) = 1$ for the whole (x, y) plane. In this case

$$\left.\begin{aligned} &\int_{-\infty}^{+\infty} \gamma(x, y)\, dx = 2 \\ \text{and} \qquad &\int_{-\infty}^{+\infty} \gamma(x, y) \cos 2u_0 x\, dx = \frac{\sin 2u_0}{u_0}. \end{aligned}\right\} \qquad \ldots\ldots(18)$$

From (3), (6) and (18) we then obtain, for this special case

$$\begin{aligned} B'(u', u_0, \infty) = 16 \int_0^1 \left[\frac{\sin^2\{(u' - u_0)\sqrt{(1 - y^2)}\}}{(u' - u_0)^2} + \frac{\sin^2\{(u' + u_0)\sqrt{(1 - y^2)}\}}{(u' + u_0)^2} \right] dy \\ + \frac{8 \sin 2u_0}{u_0(u'^2 - u_0^2)} \int_0^1 [\cos\{2u_0\sqrt{(1 - y^2)}\} - \cos\{2u'\sqrt{(1 - y^2)}\}]\, dy \end{aligned} \qquad \ldots\ldots(19)$$

which, by use of the substitution $y = \cos\theta$, leads to the result

$$B'(u', u_0, \infty) = 4\pi \left\{ \frac{H_1(2u' - 2u_0)}{(u' - u_0)^2} + \frac{H_1(2u' + 2u_0)}{(u' + u_0)^2} + \frac{\sin 2u_0}{u_0} \frac{H_1(2u') - H_1(2u_0)}{u'^2 - u_0^2} \right\} \qquad \ldots\ldots(20)$$

where $H_1(u)$ is the Struve function of unit order (Watson 1942). When $u_0 = 0$, (20) simplifies to

$$B'(u', 0, \infty) = 16\pi \frac{H_1(2u')}{u'^2} \qquad \ldots\ldots(21)$$

which is the known result for a single incoherent line source.

§ 3. NUMERICAL RESULTS

The intensity distributions along the axis of u' were computed for $u_0 = 1{\cdot}5$, 2, 2·5; $\sigma = 0$, 0·5, 1 and ∞, and in some cases, also for intermediate values. For each value of σ, the ratio of condenser to objective aperture, the value of u_0 which produced a decrease of 20% in intensity at the mid-point of the image was determined by graphical interpolation. The corresponding value of K in the formula $\xi = K\lambda/n \sin\alpha$ was then found from the relation

$$u_0 = \frac{2\pi}{\lambda}(n \sin\alpha)\left(\frac{\xi}{2}\right) = \pi K.$$

In figures 1–4 are shown curves of the intensity distribution in the images of pairs of transparent lines when the ratio of condenser aperture to that of the objective has the values $\sigma \to 0$, $\sigma = 0{\cdot}5$, $\sigma = 1$, and $\sigma \to \infty$, and for separations $2u_0$ given by $u_0 = 1{\cdot}5$, 2 and 2·5. The curve in figure 5 shows the variation of K with σ. The broken line is the corresponding curve for the resolution of two point sources given by Hopkins and Barham (1950).

Values of K in the formula $\xi = K\lambda/n \sin\alpha$ for Varying Conditions of Illumination

σ		0	1	∞
Lateral slit illumination	2 lines	0·66	0·45	
Illumination by circular source	2 points	0·84	0·61	0·61
	2 lines	0·71	0·60	0·53

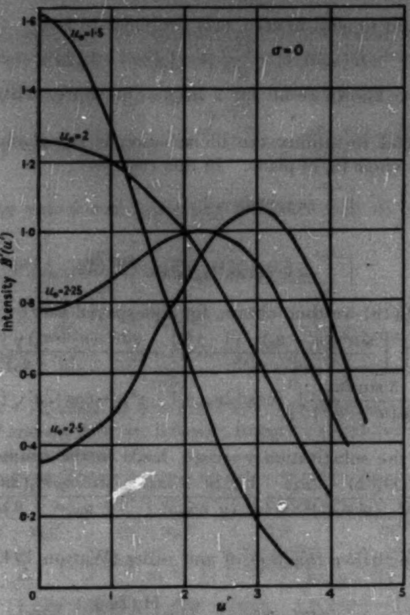


Figure 1.

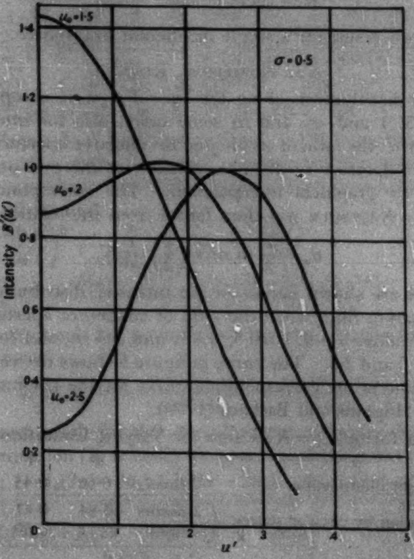


Figure 2.

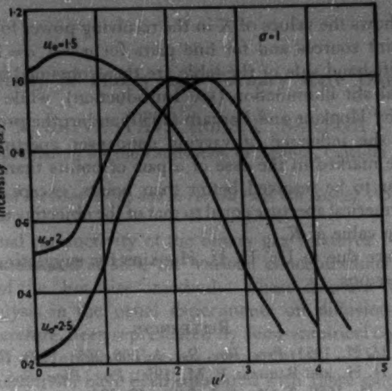


Figure 3.

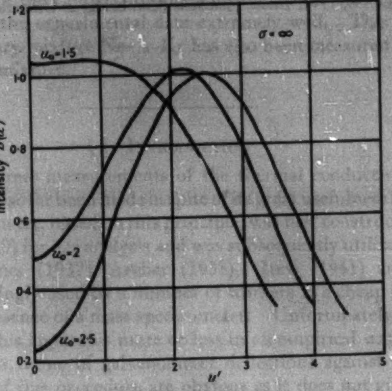


Figure 4.

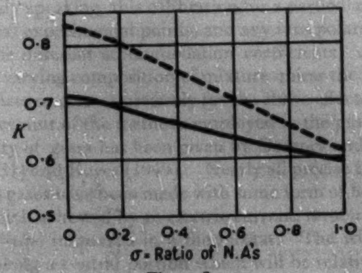


Figure 5.

The table shows the values of K in the resolving power formula $\xi = K\lambda/n \sin \alpha$ for pairs of point sources and for line pairs for $\sigma \rightarrow 0$, $\sigma = 1$ and $\sigma \rightarrow \infty$. The values in the left-hand side of the table are those obtained by Osterberg (1950) assuming lateral slit illumination (see Introduction), while those on the right were obtained by Hopkins and Barham (1950) and in the present paper. It will be noted that the influence of varying condenser aperture on resolution is somewhat more marked in the case of a pair of points than for a pair of lines; but lines appear to be resolved better than points, except when $\sigma = 1$. For a condenser of numerical aperture equal to that of the objective there is no significant difference in the value of K.

My thanks are due to Dr. H. H. Hopkins for suggesting this problem and for useful discussions.

REFERENCES

HOPKINS, H. H., 1951, *Proc. Roy. Soc.* A, **208**, 263; 1953, *Ibid.*, **217**, 408.
HOPKINS, H. H., and BARHAM, P. M., 1950, *Proc. Phys. Soc.* B, **63**, 737.
OSTERBERG, H., 1950, *J. Opt. Soc. Amer.*, **40**, 295.
SPARROW, C. M., 1916, *Astrophys. J.*, **44**, 76.
WATSON, G. N., 1942, *Bessel Functions* (Cambridge : University Press).

Thermal Conductivity of Binary and Ternary Rare Gas Mixtures

By B. N. SRIVASTAVA and S. C. SAXENA
Indian Association for the Cultivation of Science, Calcutta

MS. received 1st May 1956, and in revised form 8th August 1956

Abstract. The thermal conductivity of the binary gas mixtures Ne–A, Ne–Kr and A–Kr has been measured at 38°c for various concentrations by using the 'thick wire' variant of the 'hot wire' method. These data are likely to be of great use for gas analysis in the usual experiments on diffusion and thermal diffusion and have, therefore, been represented by semi-empirical equations giving the conductivity over the entire range of concentration. The experimental values of thermal conductivity have been compared with those calculated on the Lennard-Jones 12-6 model and the modified Buckingham exp-six model. A simple modification of the Lindsay–Bromley formula has been suggested and is found to reproduce the experimental data extremely well. The thermal conductivity of the ternary mixture Ne–A–Kr has also been measured for different compositions of the mixture.

§ 1. Introduction

Very few accurate measurements of the thermal conductivity of gaseous mixtures have so far been made in spite of its great usefulness in gas analysis. The katharometer, based on this principle, was first constructed by Daynes and Shakespear (1920) for gas analysis and was subsequently utilized by Ibbs and Hirst (1929), Trenner (1937), Archer (1938), Grew (1941) and others. It is now being increasingly used by a number of workers as a cheap instrument for gas analysis in the absence of a mass spectrometer. Unfortunately, most of these workers have used this apparatus more or less in an empirical way and obtained calibration curves in terms of galvanometer deflections against concentration. The disadvantages of this procedure are obvious as it does not give directly the thermal conductivity of the mixture though the latter depends upon the composition alone. Rigorously speaking, this calibration for a particular gas mixture is correct only at the actual experimental points, and any interpolation is subject to errors arising out of the different accommodation coefficients, wall effects and convection effects with varying composition of mixture unless the katharometer is so designed as to eliminate or correct precisely for the above effects.

A comprehensive account of the methods employed in the past for measuring the thermal conductivity of gases has been given by Gregory and Archer (1926), Trautz and Zundel (1931) and Keyes (1949). Nearly all precise thermal conductivity measurements on gases have been made with some form of hot wire method. In this a metal wire, which is heated by an electric current, is mounted axially in a vertical glass or metal tube immersed in a thermostat. The wire remains at a uniform temperature along a central portion which will be relatively greater the longer the wire and the smaller its diameter. The lateral heat transfer through the gas is limited only to a central portion of the tube and the principal correction to be applied is the heat conducted longitudinally by the wire. This difficulty has been met by applying three different types of procedure, (1) the potential lead method, (2) the compensating cell method and (3) the thick wire method. In the last, developed by Kannuluik and Martin (1933), and Kannuluik and Carman (1952), the end effect is rather large but is corrected almost exactly by a rigorous application of the theory. In view of its great simplicity this method has been chosen by us for constructing a universal type of katharometer which may be used for gas analysis in the various experiments on transport properties planned in this laboratory.

With this end in view Srivastava and Madan (1954) constructed a conductivity cell based on this method but used a relatively thinner wire and measured the thermal conductivity of A–Ne mixture. The cell design has been further modified in the present investigations and an all-metal cell developed in place of their glass cell. Further, a comparatively thicker wire has been used which not only reduces the temperature jump effect but increases the mechanical stability of the apparatus also.

§ 2. THEORY AND FORMULAE

A detailed account of the theory of the hot wire method as used in the present work was given by Kannuluik and Martin (1934) which takes into account the actual distribution of temperature in the wire and the ambient gas, subject to the boundary condition that both ends of the wire and the wall of the tube are always kept at the same temperature, which, for convenience, is taken as an arbitrary zero. However, as shown by them, it is more convenient to assume a strictly radial flow of heat from the wire surface and thereafter to apply a correction for the slight axial flow at the ends. With the assumption of radial flow the differential equation for the flow of heat is

$$\pi r_1^2 \lambda \frac{d^2\theta}{dx^2} + 2\pi r_1 h\theta + \frac{I^2\rho_0(1+\alpha\theta)}{J} = 0 \qquad \ldots\ldots(1)$$

where

$$h = K_u / r_1 \ln (r_2/r_1) \qquad \ldots\ldots(2)$$

and λ, K_u are respectively the thermal conductivities of the wire and the gas, r_1, r_2 the radius of the wire and internal radius of the tube respectively, ρ_0 the resistance per unit length of the wire at 0°c, α the temperature coefficient of resistance of the wire and J the mechanical equivalent of heat. The solution of (1) is

$$f(\beta l) = \frac{1}{(\beta l)^2}\left(1 - \frac{\tanh \beta l}{\beta l}\right) = \frac{2\pi r_1^2 \lambda J (R-R_0)}{R_0^2 I^2 \alpha l}, \qquad \ldots\ldots(3)$$

with

$$\beta^2 = \frac{2h}{r_1\lambda} - \frac{I^2 R_0 \alpha}{2\pi r_1^2 \lambda J l},$$

where R is the resistance of the wire when a current I amperes is flowing, R_0 is the resistance at the bath temperature, i.e. for $I=0$, $2l$ is the length of the cell wire. However, the exact solution is given by

$$\pi(R-R_0)/2R_0\alpha = C_1 N_{01} - \tfrac{1}{3} C_3 N_{03} + \tfrac{1}{5} C_5 N_{05} \qquad \ldots\ldots(4)$$

with

$$C_n = \pm 4I^2\rho_0/n\pi J \div [\{\pi r_1^2 \lambda (n\pi/2l)^2 - I^2\rho_0\alpha/J\} N_{0n} - 2\pi r_1 K (n\pi/2l) N_{13}]$$

where N_{0n} and N_{1n} involve Bessel functions $I_n(z)$ and $K_0(z)$ and n has suitable integral values. The radial flow correction which is of the order of one per cent is the difference in the conductivity values obtained from equations (2) and (4).

If the hot wire cell is highly evacuated the conduction loss through the gas is replaced by the radiation loss h_r and will be in general small so that β will also be small. Under such conditions equation (3) can be expanded out to yield

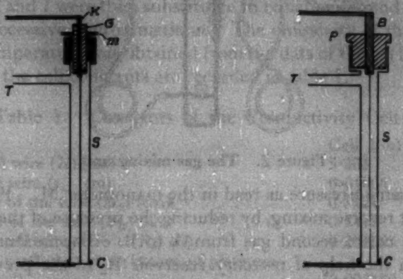
$$\lambda = \frac{1}{6}\frac{RR_0I^2\alpha l}{\pi r_1^2 J(R-R_0)}\left(1+\frac{1}{30}\frac{RI^2\alpha l}{\pi r_1^2\lambda J}\right)\left(1-\frac{4}{5}\frac{h_r l^2}{r_1\lambda}\right) \qquad \ldots\ldots(5)$$

which is useful for determining the wire conductivity at the temperature in question.

§ 3. Description of the Apparatus

3.1. *The Conductivity Cell*

The conductivity cell forms a part of the conductivity analyser, the design of the latter being substantially similar to that of Grew (1941). The conductivity cell was, however, carefully designed to be convection-free and as strictly satisfying the boundary conditions assumed in § 2 as practicable. It was made out of a stainless steel rod of 1 cm diameter and was properly shaped as shown in figure 1 by mounting it on an accurate lathe. Two designs of the cell were used, one using a commercial metal–glass seal (shown in 1 (*a*)) and the other using a Perspex rod (shown in 1 (*b*)). In both a hole was bored throughout its entire length which was uniform to 1/1000 of an inch over the whole effective length of the conductivity cell, the wall thickness being approximately 1 mm. In cell 1 (*a*) the outer metal cover m of the glass–metal seal was first made accurately cylindrical and then with respect to this an accurately centred hole was bored in the inner metal rod K in which platinum wire of diameter 0·2 mm was carefully soldered so as to be accurately axial with the outer metal cylinder S. This wire passed through an accurately centred hole in the copper cap C fitted at the other end of the cell. The wire was then made taut and carefully soldered to the cap. In the second design of the cell a tightly fitting cylindrical plug P was prepared by mounting a Perspex rod on the lathe and then a brass rod B of 3 mm diameter was inserted. In this brass rod an accurately centred hole was made by mounting on the lathe and platinum wire soldered as before. The Perspex plug was fitted vacuum tight into the cylinder



(*a*) (*b*)

Figure 1. The conductivity cell.

by means of a solution of Perspex and chloroform and a coating of glyptol. A side tube T of brass was soldered to the conductivity cells and connected to the conductivity analyser by means of a graded seal in case of cell 1 (*a*) and an araldite seal in cell 1 (*b*). Actual experiments showed that the design 1 (*b*) was a little better than the design 1 (*a*) but the latter will be more useful at high temperatures.

3.2. *The Gas Mixing Unit*

The gas mixing unit is shown in figure 2 and is self-explanatory. Suitable quantities of the desired gases were successively taken in the small pipette A and transferred with the help of mercury reservoir R_1 into the burette B. The concentration of the mixture in B was determined by observing the respective volumes at the same pressure as read in the manometer M. Particular care was taken to prevent reverse mixing, by reducing the pressure of the first gas in B and transferring the entire second gas from A to B. The mixture so formed was transferred with the help of mercury reservoir R_2 to the previously evacuated conductivity analyser.

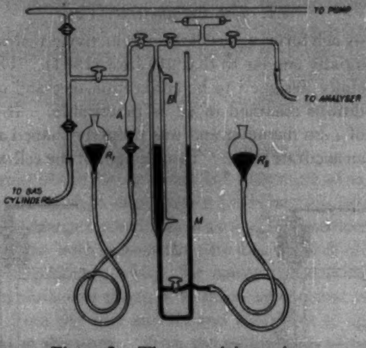


Figure 2. The gas mixing unit.

3.3. *The Thermostat and the Pressure Regulation Unit*

The entire conductivity analyser together with the cell was completely immersed in a large thermostatic bath filled with transformer oil and fitted with a glass side to enable observation from outside. Efficient stirring in conjunction with a mercury–ether regulator and an electromagnetic relay enabled the bath temperature to be kept constant to 0·05°c. To evacuate the conductivity analyser as well as to compress the experimental gas in the conductivity cell to a fixed pressure a specially designed pressure regulation system was employed.

3.4. *Electrical Circuit*

The resistance of the cell wire being only 0·2 ohm, the potentiometer method was used to measure the small change in its resistance for different compositions of the mixture. A series circuit was formed with a standard resistance, a variable resistor and a high-capacity battery. In order to obtain a constant current from the battery it was closed through a bleeder resistance when not in use. A Pye precision vernier potentiometer was used reading up to 1 microvolt with the absolute accuracy to ±0·004% or ±1 μv, but the relative accuracy of the two readings would be greater. A standard one-ohm resistance of manganin manufactured by the Cambridge Instrument Co. was used.

§ 4. CALIBRATION OF THE CONDUCTIVITY CELL

The geometrical constants of the cell, viz. the length and diameter of the cell wire and the internal radius of the tube, were first determined accurately. For determining R_0 the conductivity cell was highly evacuated and the resistance R of the cell wire was measured for different currents I. Theory shows that a linear relation connects $1/R$ and I^2 which were, therefore, plotted and R_0 was obtained by extrapolation to zero current. For determining α at the bath temperature the resistance of the cell wire was measured at ice, room and steam temperatures and these values were then substituted in the relation $R_t = R_0(1 + At + Bt^2)$, and the constants A and B determined. α was then determined by the defining relation $\alpha = (1/R)(dR/dT)$ at the temperature in question. This value of α was checked and confirmed several times by finding the resistance of the cell wire at several bath temperatures.

To determine λ the cell was highly evacuated by an oil diffusion pump and the resistance of the cell wire R measured for several values of the heating current I. These values of R and I were then substituted in equation (5) and λ determined by the method of successive approximations. The emissivity of the platinum wire at the required temperature was obtained from the data of Geiss (1925).

The values of the cell constants are recorded in table 1.

Table 1. Constants of the Conductivity Cell

	Cell 1 (*a*)	Cell 1 (*b*)
Length of the cell wire ($2l$) (cm)	5·201	4·802
Radius of the cell wire (r_1) (cm)	0·01006	0·01006
Internal diameter of the tube ($2r_2$) (cm)	0·3124	0·3036
External diameter of the tube ($2r_3$) (cm)	0·5410	0·5969
Resistance of the cell wire at 38°c (R_0) (Ω)	0·20568	0·18987
The temperature coefficient of resistance of platinum wire (α) (deg^{-1})	0·003150	0·003150
The mean thermal conductivity of the platinum wire (λ) ($\text{cal cm}^{-1}\,\text{sec}^{-1}\,\text{deg}^{-1}$)	0·168	0·168
The cell constant $(1-C)$	0·9986	0·9982

§ 5. EXPERIMENTAL RESULTS

The only experimental data on binary gas mixtures of rare gases are those of Wachsmuth (1908) on He–A and Srivastava and Madan (1954) on Ne–A while none exists on ternary mixtures. There are, however, a few data on binary mixtures of diatomic and polyatomic gases by Weber (1917), Wassiljewa (1904), Ibbs and Hirst (1929), Keyes (1951, 1952) and Bennett and Vines (1955). In the present paper we have found the thermal conductivity of the binary mixtures Ne–A, Ne–Kr, A–Kr and the ternary mixture Ne–A–Kr. The gases used were supplied by the British Oxygen Co., England; neon, argon were quoted as spectroscopically pure while krypton contained traces of xenon.

A series of observations of R for the same heating current I taken at different gas pressures from 3 to 20 cm of mercury showed that conductivities above 6 cm of mercury were independent of the pressure, indicating a complete absence of convection and negligible temperature jump effect above this pressure. We have therefore taken the observations at a pressure of 10 cm of mercury. For the sake of illustration the values of I and $R-R_0$ for the binary mixtures of Ne–A are given in table 2 for various compositions and are merely representative of a set of readings. The apparent conductivities K_u of the gas mixtures were calculated from equations (3) and (2). The conductivity was then reduced to the bath temperature and also corrected for radial flow and radiation and wall effects to give the conductivity K'. To correct for any possible asymmetry of the cell, the cell constant C occurring in the relation $K=K'(1-C)$ was determined by assuming as standard the conductivity of argon given by Kannuluik and Carman (1952). The values of conductivity K after applying all corrections are given in the last column. The results for Ne–Kr and A–Kr mixtures are recorded in table 3. All these results are plotted in figure 3. All the three plots are seen to depart from a straight line, having in general a concavity upwards, which is obviously due to unlike interactions playing a significant part.

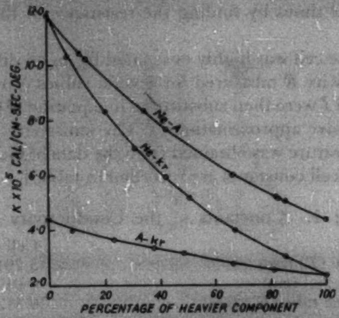


Figure 3. Conductivity of gas mixtures of different composition.

Table 2. Ne–A Mixture

(1)	(2)	(3)	(4)	(5)	(6)
			Cell 1 (*b*)		
0	336·63	2·11	12·094	11·847	11·83
11·83	348·83	2·48	10·680	10·491	10·47
13·70	349·20	2·54	10·443	10·266	10·25
34·72	348·35	3·05	8·472	8·304	8·30
66·88	343·29	3·88	6·216	6·082	6·07
82·86	349·29	4·60	5·310	5·191	5·15
86·60	348·81	4·70	5·159	5·042	5·03
100·00	226·37	2·20	4·467	4·391	
100·00	335·15	4·86	4·497	4·394	4·38
			Cell 1 (*a*)		
0	339·37	1·93	12·048	11·801	11·78
31·24	336·00	2·56	8·570	8·403	8·39
42·15	335·57	2·76	7·853	7·695	7·68
83·81	338·27	3·98	5·279	5·164	5·16
100·00	266·46	2·77	4·475	4·391	4·38

(1) Percentage of the heavier component, (2) I (mA), (3) $R-R_0$ (mA), (4) $K_u \times 10^5$, (5) $K' \times 10^5$, (6) $K \times 10^5$, (4)–(6) in (cal cm^{-1} sec^{-1} deg^{-1}).

Table 3. Ne–Kr and A–Kr Mixtures

(1)	(2)	(3)	(4)	(5)	(6)
Cell 1 (*b*)					
30·92	7·125	6·96	23·38	3·766	3·66
50·70	5·286	5·16	37·95	3·435	3·34
67·07	4·093	3·98	66·83	2·866	2·77
85·56	3·153	3·07	81·15	2·640	2·56
100·0	2·415	2·34			
Cell 1 (*a*)					
7·12	10·544	10·35	8·66	4·083	3·98
20·76	8·493	8·31	48·40	3·250	3·17
42·77	6·049	5·91			
100·0	2·399	2·33			

Ne–Kr: (1) percentage of the heavier component, (2) $K_u \times 10^5$, (3) $K \times 10^5$ (cal cm^{-1} sec^{-1} deg^{-1}); A–Kr: (4) percentage of the heavier component, (5) $K_u \times 10^5$, (6) $K \times 10^5$. (2), (3), (5) and (6) in (cal cm^{-1} sec^{-1} deg^{-1}).

§ 6. EXPERIMENTS WITH TERNARY GAS MIXTURE

The experimental data on the ternary gas mixture of neon, argon and krypton are recorded in table 4 for different compositions of the mixture. Each mixture was prepared afresh so that any possible errors in the composition of one mixture may not be carried on to the next. No fixed order of mixing the gases was adopted in order to avoid possibility of any systematic errors. Observations were taken with both the cells but those with cell 1 (*b*) are believed to be more accurate.

Table 4

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cell 1 (*b*)							
13·87	71·72	14·41	337·45	4·83	4·614	4·508	4·50
18·61	14·49	66·90	335·88	5·90	3·573	3·479	3·47
30·19	38·48	31·33	337·86	4·59	4·917	4·808	4·80
45·37	15·67	38·96	336·07	4·07	5·576	5·457	5·45
59·84	13·01	27·15	335·86	3·46	6·731	6·593	6·58
Cell 1 (*a*)							
12·79	15·69	71·52	336·02	5·62	3·334	3·245	3·24
79·19	14·17	6·64	338·18	2·58	8·635	8·468	8·45

(1) Percentage of neon, (2) percentage of argon, (3) percentage of krypton, (4) I (mA), (5) $R-R_0$ (mΩ), (6) $K_u \times 10^5$ (cal cm^{-1} sec^{-1} deg^{-1}), (7) $K' \times 10^5$ (cal cm^{-1} sec^{-1} deg^{-1}), (8) $K \times 10^5$ (cal cm^{-1} sec^{-1} deg^{-1}).

§ 7. COMPARISON OF THEORY AND EXPERIMENT

On Chapman–Enskog theory the thermal conductivity of a binary mixture of monatomic gases $[K_{\mathrm{mix}}]_1$ can be written in the form

$$[K_{\mathrm{mix}}] = (1+Z)/(X+Y). \qquad \ldots\ldots(6)$$

The expressions for X, Y and Z have been given by Hirschfelder, Curtiss and Bird (1954) in a form suitable for computation, and these have been utilized to calculate the thermal conductivity of the binary mixtures. For theoretical calculation the Lennard-Jones 12-6 model

$$E_{ij}(r) = 4\epsilon_{ij}\left[\left(\frac{r_{ij}}{r}\right)^{12} - \left(\frac{r_{ij}}{r}\right)^{6}\right],$$

and the modified Buckingham exp-six model

$$E(r) = \frac{\epsilon}{1-6/\alpha}\left[\frac{6}{\alpha}\exp\left[\alpha(1-r/r_m)\right] - \left(\frac{r_m}{r}\right)^{6}\right]$$

have been utilized, where the terms have their usual meaning. For like interactions, values of the potential parameters given by Hirschfelder, Bird and Spotz (1948) for the 12-6 model and by Mason and Rice (1954 b) for the exp-six model were used. For krypton, however, the values given by Srivastava and Srivastava (1957) for $\alpha = 13{\cdot}1$, $\epsilon/k = 135{\cdot}5$°K and $r_m = 4{\cdot}341$ Å have been used. For unlike interaction, the values given by Saxena (1955) were utilized except for A–Kr, for which $\epsilon_{12}/k = 160{\cdot}5$°K and $r_{12} = 3{\cdot}485$ Å have been used. For the exp-six model the values for unlike interactions were taken from Srivastava and Srivastava (1957) except for A–Kr, for which the combination rules of Mason and Rice (1954 a) have been

Table 5. Ne–A Mixture

(1)	(2)	(3)	(4)	(5)
0	11·83	11·83	11·81	11·83
10	10·63	10·64	10·62	10·65
20	9·60	9·60	9·59	9·61
30	8·67	8·68	8·67	8·69
40	7·85	7·86	7·85	7·88
50	7·12	7·12	7·11	7·14
60	6·45	6·46	6·45	6·48
70	5·86	5·85	5·85	5·88
80	5·32	5·31	5·30	5·33
90	4·82	4·81	4·80	4·84
100	4·38	4·35	4·34	4·38

(1) Percentage of the heavier component, (2) observed values of $K \times 10^5$, (3)–(5) calculated values of $K \times 10^5$ by (3) 12-6, (4) exp-six, (5) modified L–B equation.

Table 6

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0	11·83	11·83	11·81	11·83	4·38	4·35	4·34	4·38
10	9·87	9·86	9·83	9·95	4·05	4·04	4·05	4·04
20	8·38	8·31	8·27	8·43	3·77	3·77	3·79	3·74
30	7·15	7·05	7·02	7·17	3·51	3·52	3·54	3·48
40	6·13	6·02	5·99	6·13	3·28	3·29	3·32	3·24
50	5·25	5·15	5·12	5·25	3·07	3·09	3·12	3·13
60	4·49	4·41	4·39	4·49	2·89	2·91	2·94	2·85
70	3·85	3·78	3·77	3·84	2·72	2·74	2·77	2·70
80	3·27	3·23	3·23	3·27	2·57	2·59	2·61	2·56
90	2·77	2·75	2·76	2·78	2·44	2·45	2·47	2·44
100	2·33	2·32	2·34	2·33	2·33	2·32	2·34	2·33

(1) Percentage of the heavier component, (2)–(5) Ne–Kr: (2) observed values of $K \times 10^5$, (3)–(5) calculated values of $K \times 10^5$ by (3) 12-6, (4) exp-six, (5) modified L–B equation; (6)–(9) A–Kr: (6) observed values of $K \times 10^5$. (7)–(9) calculated values of $K \times 10^5$ by (7) 12-6, (8) exp-six, (9) modified L–B equation.

assumed. For convenience of reference the experimental data at 38°C were plotted on a large graph paper and read at regular intervals of composition. These are recorded in column 2 of table 5 for Ne–A mixture, along with the theoretically calculated values on the 12-6 and exp-six models. In table 6 the data for Ne–Kr and A–Kr mixtures are recorded. The agreement between theory and experiment is seen to be very close showing a simultaneous check of the theory as well as the accuracy of the potential parameters.

§ 8. DISCUSSION OF RESULTS

The rigorous formula for the thermal conductivity of a binary mixture expressed by equation (6) is very complicated and is not at all handy for a rapid determination of the composition of a mixture from a measurement of thermal conductivity as is often required in gas analysis. For this purpose an empirical equation connecting the conductivity with the composition might be more helpful. After trials it was found that an equation of the type originally suggested by Wassiljewa (1904)

$$K_{\mathrm{mix}}=\frac{K_1}{1+A_{12}(x_2/x_1)}+\frac{K_2}{1+A_{21}(x_1/x_2)} \qquad \ldots\ldots(7)$$

reproduces the experimental data over the entire range within 1% if the constants A_{12} and A_{21} are properly chosen. From our experimental data we have found the values of A_{12} and A_{21} respectively to be as follows: Ne–A mixture, $A_{12}=0{\cdot}597$, $A_{21}=1{\cdot}690$; Ne–Kr mixture, $A_{12}=0{\cdot}431$, $A_{21}=2{\cdot}303$; A–Kr mixture, $A_{12}=0{\cdot}626$, $A_{21}=1{\cdot}627$.

Adopting an older scheme of Sutherland (1895) for correlating mixtures, Lindsay and Bromley (1950) have given an empirical formula for binary thermal conductivity of the form (7) where A_{12} is given by

$$A_{12}=\frac{1}{4}\left[1+\left\{\frac{\eta_1}{\eta_2}\left(\frac{M_2}{M_1}\right)^{3/4}\frac{1+S_1/T}{1+S_2/T}\right\}^{1/2}\right]^2\frac{1+S_{12}/T}{1+S_1/T}, \qquad \ldots\ldots(8)$$

and η_1, η_2 and S_1, S_2 are the viscosities and Sutherland constants of the two components respectively. S_{12} is assumed to be the geometric mean of S_1 and S_2. A_{21} is obtained from A_{12} by interchanging the subscripts. Actual calculation using equations (7) and (8) shows a discrepancy of the order of 5% between the calculated and observed values. From a close examination of Enskog's formula for a binary mixture we suggest the following modification to the Lindsay–Bromley formula:

$$K_{\mathrm{mix}}=\frac{K_1}{1+A_{12}(x_2/x_1)}+\frac{K_2}{1+A_{21}(x_1/x_2)}+\frac{C(K_1K_2)^{1/2}}{\{1+A_{12}(x_2/x_1)\}\{1+A_{21}(x_1/x_2)\}} \qquad \ldots\ldots(9)$$

where C is a constant which can be calculated if the conductivity at one composition is known. Values of this constant C for the three gas pairs Ne–A, Ne–Kr and A–Kr are $-0{\cdot}0937$, $-0{\cdot}221$ and $-0{\cdot}245$ respectively. Values of the thermal conductivity calculated from this modified Lindsay–Bromley formula are recorded in the last column of table 5 and in columns (5) and (9) of table 6. It will be seen that the modified Lindsay–Bromley formula reproduces the observed data to a high degree of accuracy.

The rigorous theory of the thermal conductivity of multicomponent gas mixtures yields a formula which is too complicated for numerical computation, even for a three-component system. In the absence of computational facilities in our laboratory we are unable to compare our experimental results with the theory. We hope, however, to get this computation done through the courtesy of the Indian Statistical Institute. The results of this computation along with a generalization of equation (7) developed by us will be published later.

ACKNOWLEDGMENTS

We record our thanks to the Council of Scientific and Industrial Research, New Delhi, for sanctioning a research project under which the present investigation has been carried out and for the award of a research scholarship to one of us (S.C.S.).

REFERENCES

ARCHER, C. T., 1938, *Proc. Roy. Soc.* A, **165**, 474.
BENNETT, L. A., and VINES, R. G., 1955, *J. Chem. Phys.*, **23**, 1587.
DAYNES, H. A., and SHAKESPEAR, G. A., 1920, *Proc. Roy. Soc.* A, **97**, 273.
GREGORY, H. S., and ARCHER, C. T., 1926, *Proc. Roy. Soc.* A, **110**, 91.
GREW, K. E., 1941, *Proc. Roy. Soc.* A, **178**, 390.
GEISS, W., 1925, *Physica*, **5**, 203.
HIRSCHFELDER, J. O., BIRD, R. B., and SPOTZ, E. L., 1948, *J. Chem. Phys.*, **16**, 968.
HIRSCHFELDER, J. O., CURTISS, C. F., and BIRD, R. B., 1954, *Molecular Theory of Gases and Liquids* (New York: John Wiley).
IBBS, T. L., and HIRST, A. A., 1929, *Proc. Roy. Soc.* A, **123**, 134.
KEYES, F. G., 1949, *Trans. Amer. Soc. Mech. Engrs*, **71**, 966; 1951, *Ibid.*, **73**, 597; 1952, *Ibid.*, **74**, 1303.
KANNULUIK, W. G., and CARMAN, E. H., 1952, *Proc. Phys. Soc.* B, **65**, 701.
KANNULUIK, W. G., and MARTIN, L. H., 1933, *Proc. Roy. Soc.* A, **141**, 144; 1934, *Ibid.*, **144**, 496.
LINDSAY, A. L., and BROMLEY, L. A., 1950, *Industr. Engng Chem.*, **42**, 1508.
MASON, E. A., and RICE, W. E., 1954 a, *J. Chem. Phys.*, **22**, 522; 1954 b, *Ibid.*, **22**, 843.
SAXENA, S. C., 1955, *Indian J. Phys.*, **29**, 131.
SRIVASTAVA, B. N., and SRIVASTAVA, K. P., 1957, *Physica*, **23**, 103.
SRIVASTAVA, B. N., and MADAN, M. P., 1954, *Proc. Nat. Inst. Sci., India*, **20**, 587.
SUTHERLAND, W., 1895, *Phil. Mag.*, **40**, 421.
TRAUTZ, M., and ZUNDEL, A., 1931, *Z. Tech. Phys.*, **12**, 273.
TRENNER, N. R., 1937, *J. Chem. Phys.*, **5**, 382.
WACHSMUTH, J., 1908, *Phys. Z.*, **9**, 235.
WASSILJEWA, A., 1904, *Phys. Z.*, **5**, 737.
WEBER, S., 1917, *Ann. Phys., Lpz.*, **54**, 325.

The Theory of Light Scattering by Suspensions of Randomly Oriented Long Prisms

By A. R. STOKES

Wheatstone Laboratory, King's College, London

Communicated by C. Domb; MS. received 14*th November* 1956, *and in final form* 22*nd December* 1956

Abstract. A general asymptotic formula is developed for the normalized intensity of scattering by long, thin prisms of specified length and cross section. This is applied to the particular case of long, thin, parallel-sided strips, both monodisperse and polydisperse. The effect of interaction between strips in contact is also considered.

§ 1. Introduction

The work described in this paper was undertaken largely in order to provide the theoretical basis for investigations by M'Ewen and Pratt (1957) on the light-scattering properties of bentonite suspensions. Consequently, the choice of the systems considered here has been partly determined by their requirements. Porod (1949) has already given some similar results, but his work does not cover all the ground needed for the present purpose, and the methods of derivation used here are different from those of Porod.

The general problem is to derive formulae for the scattered intensity at angle θ, averaged over all orientations of the scattering particles, and normalized so as to be unity at $\theta=0$. This is the quantity obtained from light-scattering measurements, and is always denoted by $P(\theta)$. For purposes of calculation, however, we use as our variable the quantity $\omega=4\pi\sin\frac{1}{2}\theta/\lambda$, and the intensity is thus a function of ω and the dimensions a, b, of the particle. We shall therefore use $\mathsf{P}(\omega)$ or $\mathsf{P}(\omega, a, b)$ to denote the normalized intensity expressed as a function of ω and the dimensions of the particle.

§ 2. General Expression for the Scattered Intensity

We use the general formula† for the scattering of electromagnetic waves by a non-absorbing particle of uniform refractive index:

$$\frac{I}{I_0}=\frac{1+\cos^2\theta}{2}\left|\frac{1}{V}\int\exp\left\{\frac{2\pi i}{\lambda}(\mathbf{s}-\mathbf{s}').\mathbf{r}\right\}d\quad\right|^2, \qquad \ldots\ldots(1)$$

where $\mathbf{r}$ is the position vector of a volume element dv in the particle, $\mathbf{s}$ and $\mathbf{s}'$ are unit vectors in the direction of propagation of the incident waves and the scattered waves respectively, V is the volume of the particle, I and I_0 are the intensities of light scattered in the directions $\mathbf{s}'$ and $\mathbf{s}$ respectively, and θ is the angle between the directions $\mathbf{s}$ and $\mathbf{s}'$.

---

† See, for instance, Guinier and Fournet (1955), Chap. 2, equations (1)–(4), from which our equation (1) may be derived. Their 2θ corresponds to our θ.

Let the particle be a prism of length a, and let us use coordinate axes (x, y, z) with the origin of coordinates at the centroid of the particle and the z-axis parallel to the length of the prism. Let the components in the directions (x, y, z) of the vector $2\pi(\mathbf{s}-\mathbf{s}')/\lambda$ be (h, k, l); then we can immediately integrate equation (1) with respect to z to give

$$\frac{I}{I_0}=\frac{1+\cos^2\theta}{2}\frac{1}{A^2}\left(\frac{\sin\frac{1}{2}al}{\frac{1}{2}al}\right)^2\left|\iint \exp i(hx+ky)\,dA\right|^2, \qquad \ldots\ldots(2)$$

where dA is an element of cross-sectional area of the prism and A is the total cross-sectional area. If dA_1 and dA_2 are two elements of cross-sectional area at points (x_1, y_1) and (x_2, y_2) respectively, then equation (2) becomes

$$\frac{I}{I_0}=\frac{1}{A^2}\frac{1+\cos^2\theta}{2}\left(\frac{\sin\frac{1}{2}al}{\frac{1}{2}al}\right)^2\iint \exp i[h(x_1-x_2)+k(y_1-y_2)]\,dA_1\,dA_2. \qquad \ldots\ldots(3)$$

We shall be concerned only with randomly oriented suspensions of particles, so we wish to average this quantity over all directions of the vector (h, k, l), keeping the length of this vector constant. This length, which we denote by ω, is given by

$$\omega=\frac{2\pi}{\lambda}|\mathbf{s}-\mathbf{s}'|=\frac{4\pi}{\lambda}\sin\tfrac{1}{2}\theta. \qquad \ldots\ldots(4)$$

The intensity of light scattered through angle θ, relative to the forward-scattered intensity, is therefore given by the average of I/I_0 over all vectors (h, k, l) whose length equals ω; that is, by the mean of I/I_0 over a sphere of radius ω in (h, k, l) space.

We introduce polar coordinates:

$$h=\omega\sin\alpha\cos\phi, \quad k=\omega\sin\alpha\sin\phi, \quad l=\omega\cos\alpha, \qquad \ldots\ldots(5)$$

so that equation (3) becomes

$$\frac{I}{I_0}=\frac{1+\cos^2\theta}{2}\frac{1}{A^2}\left\{\frac{\sin(\frac{1}{2}a\omega\cos\alpha)}{\frac{1}{2}a\omega\cos\alpha}\right\}^2 \times\iint \exp\{i\omega\sin\alpha[(x_1-x_2)\cos\phi+(y_1-y_2)\sin\phi]\}\,dA_1\,dA_2. \qquad \ldots\ldots(6)$$

The mean of I/I_0 over all directions, divided by the factor $(1+\cos^2\theta)/2$, is denoted by $P(\theta)$ and is given by:

$$P(\theta)=\frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi}\frac{I}{I_0}\frac{2}{1+\cos^2\theta}\sin\alpha\,d\phi\,d\alpha. \qquad \ldots\ldots(7)$$

If in equation (6) we put $x_1-x_2=r_{12}\cos\eta$; $y_1-y_2=r_{12}\sin\eta$, where r_{12} is the distance between the elements of area dA_1 and dA_2, then, by combining equations (6) and (7), and integrating with respect to ϕ, we obtain

$$P(\theta)=\mathsf{P}(\omega)=\frac{1}{2}\frac{1}{A^2}\int_0^{\pi}\left\{\frac{\sin(\frac{1}{2}a\omega\cos\alpha)}{\frac{1}{2}a\omega\cos\alpha}\right\}^2\iint J_0(\omega r_{12}\sin\alpha)\sin\alpha\,dA_1\,dA_2\,d\alpha. \qquad \ldots\ldots(8)$$

We shall denote $\dfrac{1}{A^2}\iint J_0(\omega r_{12}\sin\alpha)\,dA_1\,dA_2$ by $\mathrm{E}(\omega\sin\alpha)$.

§ 3. FIRST APPROXIMATION FOR VERY LONG PRISMS

If $a\omega$ is very large, then $\{\sin(\frac{1}{2}a\omega\cos\alpha)/\frac{1}{2}a\omega\cos\alpha\}^2$ behaves as a peak function of $\cos\alpha$, being negligibly small except when $\cos\alpha$ is very near zero, and its integral with respect to $\cos\alpha$ being equal to $\pi/\frac{1}{2}a\omega$. It follows that the integral with respect to α of equation (8) involves only the value of $E(\omega\cos\alpha)$ at $\alpha=\pi/2$, and is:

$$\mathsf{P}(\omega)=\frac{1}{2}\,\frac{2\pi}{a\omega}\,\frac{1}{A^2}\iint J_0(\omega r_{12})\,dA_1\,dA_2=\frac{\pi}{a\omega}\,E(\omega). \qquad \ldots\ldots(9)$$

This formula is clearly analogous to the well-known formula of Debye (1915) (see also Guinier and Fournet 1955, Chap. 2, equation (8)):

$$\mathsf{P}(\omega)=\frac{1}{V^2}\iint\frac{\sin\omega r_{12}}{\omega r_{12}}\,dv_1\,dv_2. \qquad \ldots\ldots(10)$$

For a long, thin strip of width b and thickness t much less than λ the expression for $E(\omega)$ reduces to

$$E(\omega)=\frac{1}{b^2}\int_0^b\int_0^b J_0[\omega(x_1-x_2)]\,dx_1\,dx_2=\frac{1}{\omega b}\int_0^{b\omega}\Lambda_1(x)\,dx \qquad \ldots\ldots(11)$$

or, in series notation:

$$E(\omega)=1-\frac{b^2\omega^2}{3.2^2.2!}+\frac{b^4\omega^4}{5.2^4.2!3!}-\frac{b^6\omega^6}{7.2^6.3!4!}+\ldots; \qquad \ldots\ldots(12)$$

and

$$P(\theta)=\mathsf{P}(\omega,a,b)=\frac{\pi}{ab\omega^2}\int_0^{b\omega}\Lambda_1(x)\,dx. \qquad \ldots\ldots(13)$$

This function is shown graphically in figure 1. The integral in equation (11) is the function $F(b\omega)$ defined generally by

$$\mathsf{P}(\omega)=P(\theta)=\frac{\pi}{ab\omega^2}F(b\omega) \qquad \ldots\ldots(13a)$$

and used by M'Ewen and Pratt (1956) in their interpretation of light-scattering measurements on bentonite. In this case it increases initially from zero and then oscillates with decreasing amplitude about the value 2·00; the beginning of this oscillation is shown in figure 5, curve (a). Consequently,

$$\mathsf{P}(\omega)\sim\frac{2\pi}{ab\omega^2}$$

for large $b\omega$.

The function $\mathsf{P}(\omega)$ in equation (13) approaches infinity instead of unity as ω approaches zero. This discrepancy is to be expected, because the formula was deduced on the assumption that $a\omega$ is large; it therefore becomes invalid when ω is small. We shall see later how the approximation for fairly small angles can be improved.

§ 4. FIRST-ORDER APPROXIMATION FOR LONG THIN STRIPS, POLYDISPERSE IN WIDTH

Consider a collection of strips such that the fraction with their widths lying between b and $b+\delta b$ is $\kappa\exp(-\kappa b)\delta b$, the length of every strip being a. Since the intensity of scattering at zero angle is proportional to b^2, the resultant intensity for this polydisperse collection is given by the following weighted mean:

$$\begin{aligned}\mathsf{P}(\omega)&=\int_0^\infty b^2\kappa e^{-\kappa b}\frac{\pi}{ab\omega^2}\int_0^{b\omega}\Lambda_1(x)\,dx\,db\Big/\!\int b^2\kappa e^{-\kappa b}\,db\\&=\frac{\pi}{a\omega}\frac{\kappa^3}{2}\int_0^\infty b^2e^{-\kappa b}\frac{1}{b\omega}\int_0^{b\omega}\Lambda_1(x)\,dx\,db \qquad \ldots\ldots(14)\\&=\frac{\pi}{a\omega}\frac{\kappa^3}{2}\int_0^\infty b^2e^{-\kappa b}\left(1-\frac{b^2\omega^2}{3.2^2.2!}+\frac{b^4\omega^4}{5.2^4.2!3!}-\frac{b^6\omega^6}{7.2^6.3!4!}+\ldots\right)db\\&=\frac{\pi}{a\omega}\left(1+\frac{\omega^2}{\kappa^2}\right)^{-1/2}. \qquad \ldots\ldots(15)\end{aligned}$$

We note that this is a much simpler expression than the corresponding expression (13) for monodisperse strips.

When ω is large

$$\mathsf{P}(\omega) \simeq \frac{\pi\kappa}{a\omega^2}$$

and if we call the weight-average breadth b_w, so that

$$b_w = \overline{b^2}/\bar{b} = 2/\kappa,$$

then

$$\mathsf{P}(\omega) \simeq \frac{2\pi}{ab_w\omega^2} \qquad \ldots\ldots(16)$$

for large ω, as for the monodisperse strips.

Now let the fraction of strips with widths lying between b and $b+\delta b$ be $\kappa^2 b \exp(-\kappa b)\,\delta b$, so as to give a distribution with a maximum. As before, we take a weighted mean of scattering intensities, and we obtain

$$\mathsf{P}(\omega) = \frac{\pi}{a\omega}\left(1+\frac{\omega^2}{\kappa^2}\right)^{-1/2}\left\{\frac{2}{3}+\frac{1}{3}\left(1+\frac{\omega^2}{\kappa^2}\right)^{-1}\right\}. \qquad \ldots\ldots(17)$$

For this distribution the weight-average breadth is

$$b_w = \overline{b^3}/\bar{b} = 3/\kappa,$$

hence

$$\mathsf{P}(\omega) \sim \frac{2\pi}{ab_w\omega^2}$$

for large ω.

Finally, if the fraction of strips with widths lying between b and $b+\delta b$ is $\frac{1}{2}\kappa^3 b^2 \exp(-\kappa b)\,\delta b$, we find, similarly, that

$$\mathsf{P}(\omega) = \frac{\pi}{a\omega}\left(1+\frac{\omega^2}{\kappa^2}\right)^{-1/2}\left\{\frac{1}{2}+\frac{1}{4}\left(1+\frac{\omega^2}{\kappa^2}\right)^{-1}+\frac{1}{4}\left(1+\frac{\omega^2}{\kappa^2}\right)^{-2}\right\}. \qquad \ldots\ldots(18)$$

The weight-average breadth b_w is $4/\kappa$ and again

$$\mathsf{P}(\omega) \sim \frac{2\pi}{ab_w\omega^2}$$

for large ω.

We assumed at first that all strips were of the same length and all were parallel-sided. This is not necessary, however, provided all are long enough for $a\omega$ to be large, and provided the statistical distribution function defines the fraction by *length* of strips that have a given width. Further, the results will hold approximately for non-uniform strips, provided that there are no appreciable changes in width of any strip over a length less than a', where a' is a length such that $a'\omega$ is large.

§ 5. IMPROVED APPROXIMATION FOR PRISMS OF FINITE LENGTH

We now evaluate the expression in equation (8) more rigorously and to a better approximation than that in equation (9). We still assume that the length a is large compared with the other dimensions of the particle, and it is therefore convenient to obtain an expression for $\mathsf{P}(\omega)$ in terms of a series of ascending powers of $(1/a)$. We shall find that the result cannot be expressed strictly as a power series in $(1/a)$, since some of the terms contain oscillating functions of a; but the *amplitude* of oscillation of the nth term is proportional to $(1/a)^n$.

We have, from equation (8),

$$\mathsf{P}(\omega)=\frac{1}{2}\int_0^{\pi}\left\{\frac{\sin(\frac{1}{2}a\omega\cos\alpha)}{\frac{1}{2}a\omega\cos\alpha}\right\}^2 \mathrm{E}(\omega\sin\alpha)\sin\alpha\,d\alpha. \qquad \ldots\ldots(19)$$

We let $\cos\alpha=u$, $\sin\alpha\,d\alpha=-du$ and $\mathrm{E}(\omega\sin\alpha)\equiv\mathrm{E}\{\omega(1-u^2)^{1/2}\}=g(u)$. Since $g(u)$ is an even function of u by this definition, we may expand it as a series:

$$g(u)=g_0+g_2u^2+g_4u^4+\ldots=\mathrm{E}(\omega\sin\alpha). \qquad \ldots\ldots(20)$$

Thus

$$\mathsf{P}(\omega)=\frac{1}{2}\int_{-1}^{1}\left\{\frac{\sin(\frac{1}{2}a\omega u)}{\frac{1}{2}a\omega u}\right\}^2(g_0+g_2u^2+g_4u^4+\ldots)\,du. \qquad \ldots\ldots(21)$$

We have to consider separately the g_0 term and the other terms in this expansion. For the g_0 term we have

$$\frac{1}{2}\int_{-1}^{1}\left\{\frac{\sin(\frac{1}{2}a\omega u)}{\frac{1}{2}a\omega u}\right\}^2 g_0\,du=g_0\left[\frac{\pi}{a\omega}-\frac{2}{a^2\omega^2}-\frac{2\sin a\omega}{a^3\omega^3}+\frac{4\cos a\omega}{a^4\omega^4}+\ldots\right], \qquad \ldots\ldots(22)$$

while for the $g_{2n}u^{2n}$ terms we have

$$\frac{1}{2}\int_{-1}^{1}\left\{\frac{\sin(\frac{1}{2}a\omega u)}{\frac{1}{2}a\omega u}\right\}^2 g_{2n}u^{2n}\,du=2g_{2n}\int_0^1\left(\frac{1}{a^2\omega^2}-\frac{\cos a\omega u}{a^2\omega^2}\right)u^{2n-2}\,du$$

$$=2g_{2n}\left[\frac{1}{2n-1}\frac{1}{a^2\omega^2}-\frac{\sin a\omega}{a^3\omega^3}-(2n-2)\frac{\cos a\omega}{a^4\omega^4}+\ldots\right]. \qquad \ldots\ldots(23)$$

We can combine (22) and (23) to give

$$\mathsf{P}(\omega)=\frac{g_0\pi}{a\,\omega}+\frac{2}{a^2\omega^2}\sum_0^\infty\frac{g_{2n}}{2n-1}-\frac{2\sin a\omega}{a^3\omega^3}\sum_0^\infty g_{2n}-\frac{2\cos a\omega}{a^4\omega^4}\sum_0^\infty(2n-2)g_{2n}+\ldots. \qquad \ldots\ldots(24)$$

But

$$g_0=g(0)=\mathrm{E}(\omega) \qquad \ldots\ldots(25)$$

and

$$\sum_0^\infty g_{2n}=g(1)=\mathrm{E}(0)=1. \qquad \ldots\ldots(26)$$

From equations (20) and (26) we may show that

$$\sum_0^\infty\frac{g_{2n}}{2n-1}=-1-\omega\int_0^{\pi/2}\mathrm{E}'(\omega\sin\alpha)\,d\alpha \qquad \ldots\ldots(27)$$

and

$$\sum_0^\infty(2n-2)g_{2n}=-\omega^2\mathrm{E}''(0)-2. \qquad \ldots\ldots(28)$$

Thus we have

$$P(\theta)=\mathsf{P}(\omega)=\frac{\pi}{a\omega}\mathrm{E}(\omega)-\frac{2}{a^2\omega^2}\left[1+\omega\int_0^{\pi/2}\mathrm{E}'(\omega\sin\alpha)\,d\alpha\right]$$

$$-\frac{2\sin a\omega}{a^3\omega^3}+\frac{2\cos a\omega}{a^4\omega^4}\{\omega^2\mathrm{E}''(0)+2\}+\ldots. \qquad \ldots\ldots(29)$$

Furthermore, from the definition of $\mathrm{E}(\omega\sin\alpha)$ following equation (8), we have

$$\int_0^{\pi/2}\omega\mathrm{E}'(\omega\sin\alpha)\,d\alpha=-\frac{1}{A^2}\iint.\int_0^{\pi/2}\omega r_{12}\mathrm{J}_1(\omega r_{12}\sin\alpha)\,d\alpha\,.\,dA_1\,dA_2$$

and if we expand the Bessel function and integrate each term with respect to α, we find that

$$1+\omega\int_0^{\pi/2} \mathrm{E}'(\omega\sin\alpha)\,d\alpha = \frac{1}{A^2}\iint \cos\omega r_{12}\,dA_1\,dA_2. \qquad \ldots\ldots(30)$$

Also $$\mathrm{E}(\omega) = \frac{1}{A^2}\iint \left(1-\tfrac{1}{6}\omega^2 r_{12}{}^2+\ldots\right)dA_1\,dA_2;$$

hence $$\omega^2\mathrm{E}''(0) = -\frac{1}{A^2}\frac{\omega^2}{2}\iint r_{12}{}^2\,dA_1\,dA_2 = -\omega^2 k^2, \qquad \ldots\ldots(30\,a)$$

where k is the radius of gyration of the prism *about the long axis passing through the centroid of the cross section.*

Thus equation (29) can now be written

$$P(\theta)=\mathbf{P}(\omega)=\frac{\pi}{a\omega}\mathrm{E}(\omega)-\frac{2}{a^2\omega^2}\frac{1}{A^2}\iint\cos\omega r_{12}\,dA_1\,dA_2$$
$$-\frac{2\sin a\omega}{a^3\omega^3}+\frac{2\cos a\omega}{a^4\omega^4}(2-\omega^2k^2)+\ldots. \qquad \ldots\ldots(31)$$

If a is large, the third, fourth and higher terms oscillate rapidly as ω increases, and their mean value over a fairly small range of values of ω is zero. A slight polydispersity in the lengths a of the particles would 'smear out' these terms and reduce their amplitude considerably; thus only the first and second terms (which do not oscillate rapidly in this way) are likely to be important.

For a thin strip, width b, this becomes

$$P(\theta)=\mathbf{P}(\omega,a,b)=\frac{\pi}{ab\omega^2}\int_0^{b\omega}\Lambda_1(x)\,dx-\frac{2}{a^2\omega^2}\left\{\frac{\sin\frac{1}{2}b\omega}{\frac{1}{2}b\omega}\right\}^2-\frac{2\sin a\omega}{a^3\omega^3}$$
$$+\frac{4\cos a\omega}{a^4\omega^4}\left(1-\frac{1}{24}b^2\omega^2\right)+\ldots. \qquad \ldots\ldots(32)$$

This formula agrees, up to the third term, with that obtained by a different method by Porod (1949, equation (4), in which his $\mu=a\omega$ and $\sigma=b\omega$), if one makes two small corrections to Porod's formula.†

For thin strips with a $\kappa\exp(-\kappa b)$ distribution of width (as in §4) we find, similarly,

$$P(\theta)=\mathbf{P}(\omega,a,\kappa)=\frac{\pi}{a\omega}\frac{\kappa}{(\kappa^2+\omega^2)^{1/2}}-\frac{2}{a^2\omega^2}\frac{\kappa^2}{\kappa^2+\omega^2}-\frac{2\sin a\omega}{a^3\omega^3}$$
$$+\frac{4\cos a\omega}{a^4\omega^4}\left(1-\frac{1}{12}\frac{\omega^2}{\kappa^2}\right)+\ldots. \qquad \ldots\ldots(33)$$

§ 6. CALCULATION OF THE P(ω) CURVE FOR THIN, RECTANGULAR STRIPS

The asymptotic formula (32) enables one to calculate $\mathbf{P}(\omega)$ for values of ω such that $a\omega$ is not too small. Thus for $a\omega$ greater than 6, the third term becomes less than 0·01, and therefore the first two terms are sufficient to give an accuracy of one per cent of the maximum scattering intensity. For small values of ω it becomes necessary to calculate directly from the formula (cf. equation (10))

$$\mathbf{P}(\omega)=\frac{1}{A^2}\iint\frac{\sin\omega r_{12}}{\omega r_{12}}\,dA_1\,dA_2,$$

---

† Replace $1/\sigma^2$ by $1/\sigma$ in the first term, and multiply the second term by $1/\mu$ (Porod, private communication).

which can be expressed in this case as

$$\begin{aligned}\mathbf{P}(\omega)&=\frac{1}{a^2b^2}\int_{-a}^{a}\int_{-b}^{b}(a-|x|)(b-|y|)\frac{\sin\omega(x^2+y^2)^{1/2}}{\omega(x^2+y^2)^{1/2}}\,dx\,dy\\&=\frac{4}{a^2b^2}\int_0^a\int_0^b(a-x)(b-y)\sum_{n=0}^{\infty}\frac{(-1)^n\omega^{2n}(x^2+y^2)^n}{(2n+1)!}\,dx\,dy\\&=4\sum_0^{\infty}b^{2n}\omega^{2n}\int_0^1\int_0^1(1-x')(1-y')\frac{(-1)^n(m^2x'^2+y'^2)^n}{(2n+1)!}\,dx'\,dy' \qquad \ldots\ldots(34)\end{aligned}$$

where $m=a/b>1$.

This is a power series in $b^2\omega^2$, the coefficients of which are obtained by performing the required integrations.

Calculations of $\mathbf{P}(\omega)$ as a function of $b\omega$ have been made from this series,[†] and in figure 1 is shown the result of these calculations for a thin strip whose length is five times its width, together with the curve obtained by taking the first two terms of (32), and the curve obtained by taking only the first term of (32), i.e. the formula of equation (13).

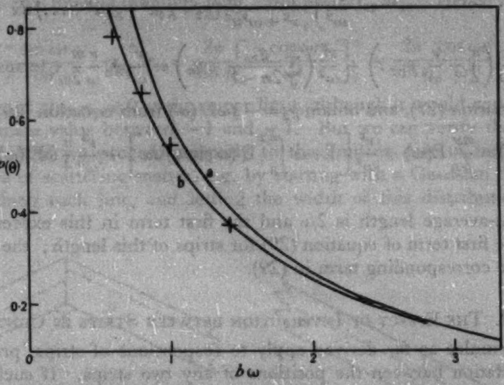


Figure 1. Curves showing normalized scattering intensity for thin strips with $a=5b$. Curve a: $P(\theta)$ as given by equation (13) or by the first term of equation (32). Curve b: $P(\theta)$ as given by the first and second terms of equation (32). The points show $P(\theta)$ as given by numerical calculation from equation (34).

The summation of the series in equation (34) becomes increasingly tedious as $b\omega$ is increased, but fortunately the expression (32) can then be used. We see in figure 1 that the curves calculated from the two series join up satisfactorily, but that the curve obtained from the first term of (32) alone deviates considerably from both the others at low values of $b\omega$.

[†] These calculations were made by Dr. M. B. M'Ewen, and I am indebted to her for the use of these results and those shown in figure 5.

§ 7. SYSTEM OF PRISMS UNIFORM IN CROSS SECTION BUT POLYDISPERSE IN LENGTH

Let the fraction of prisms with lengths lying between a and $a+\delta a$ be $\nu \exp(-\nu a)\,\delta a$. If we take the weighted mean of $\mathbf{P}(\omega)$ as given by equation (20) we obtain

$$\mathbf{P}(\omega)=\frac{\frac{1}{2}\int_0^\infty\int_{-1}^{1} \nu a^2 \exp(-\nu a)\{\sin \tfrac{1}{2}a\omega u/\tfrac{1}{2}a\omega u\}^2(g_0+g_2u^2+g_4u^4+\ldots)\,du\,da}{\int_0^\infty \nu a^2 \exp(-\nu a)\,da}.$$

On performing the integrations with respect to a we have

$$\begin{aligned}\mathbf{P}(\omega) &= \int_0^1 \frac{\nu^2}{\nu^2+\omega^2u^2}(g_0+g_2u^2+g_4u^4+\ldots)\,du\\ &=g_0\frac{\nu}{\omega}\tan^{-1}\frac{\omega}{\nu}+\int_0^1\frac{\nu^2u^2}{\nu^2+\omega^2u^2}(g_2+g_4u^2+g_6u^4+\ldots)\,du\\ &=g_0\frac{\nu}{\omega}\left(\frac{\pi}{2}-\tan^{-1}\frac{\nu}{\omega}\right)+\int_0^1\frac{\nu^2}{\omega^2}(g_2+g_4u^2+g_6u^4+\ldots)\,du\\ &\qquad -\frac{\nu^2}{\omega^2}\int_0^1\frac{\nu^2}{\nu^2+\omega^2u^2}(g_2+g_4u^2+g_6u^4+\ldots)\,du\\ &=g_0\frac{\nu}{\omega}\left(\frac{\pi}{2}-\frac{\nu}{\omega}+\ldots\right)+\frac{\nu^2}{\omega^2}\left(\sum_0^\infty\frac{g_{2n}}{2n-1}+g_0\right)-\frac{\nu^2}{\omega^2}g_2\frac{\nu}{\omega}\frac{\pi}{2}+\text{terms in }\frac{\nu^4}{\omega^4}\text{ etc.}\end{aligned}$$

We use equation (27), and obtain $g_2=-\frac{1}{2}\omega\mathrm{E}'(\omega)$ from equation (20); hence

$$P(\theta)=\mathbf{P}(\omega)=\frac{\pi\nu}{2\omega}\mathrm{E}(\omega)-\frac{\nu^2}{\omega^2}\left[1+\omega\int_0^{\pi/2}\mathrm{E}'(\omega\sin\alpha)\,d\alpha\right]+\frac{\nu^3}{\omega^3}\frac{\pi}{4}\omega\mathrm{E}'(\omega)+\ldots. \qquad \ldots\ldots(35)$$

The weight-average length is $2/\nu$ and the first term in this expression is thus equal to the first term of equation (29) for strips of this length; the second term is twice the corresponding term in (29).

§ 8. THE EFFECT OF INTERACTION BETWEEN STRIPS IN CONTACT

The formulae so far derived apply to suspensions of strips, provided there is no correlation between the positions of any two strips. If such correlation occurs, it will affect the scattered intensity, especially at lower angles.

In order to calculate this effect, we apply the formula of Debye:

$$\mathbf{P}(\omega)=\frac{1}{V^2}\iint\frac{\sin\omega r_{12}}{\omega r_{12}}\,dv_1\,dv_2, \qquad \ldots\ldots(10)$$

to a pair of strips, keeping V as the volume of *one* particle. We divide the integral into four parts, namely: (*a*) that for which dv_1 and dv_2 are both within the first strip, (*b*) that for which dv_1 and dv_2 are both in the second strip, and (*c*) and (*d*) those for which one element is in one strip and the other element is in the other. The parts (*a*) and (*b*) are just the normalized intensities for the two particles, and (*c*) and (*d*) therefore represent the effect of the interaction; thus the interaction effect *per strip* is represented by (*c*) *or* (*d*) alone, that is, by the integral (10) in which all dv_1 are in the first strip and all dv_2 are in the second.

As the strips are thin, we can replace elements of volume in the integral (10) by elements of area. It is convenient to start the integration of (10) by dividing each strip into line elements parallel to its edge, and integrating along these lines. This means that we must first evaluate the integral (10) along a pair of lines which are generally skew to each other.

Let O_1 and O_2 be the points on the two lines in which their common perpendicular cuts them, and let $O_1O_2=d$. Let ψ be the angle between the lines. If dx_1 and dx_2 are two elements of length of the lines at distance x_1 and x_2 from O_1 and O_2 respectively, then the required integral is

$$\iint \frac{\sin \omega r_{12}}{\omega r_{12}}\, dx_1\, dx_2$$

where

$$r_{12}^2 = x_1^2 + x_2^2 - 2x_1x_2 \cos\psi + d^2 = \rho^2 + d^2$$

where ρ is the distance from the origin of a point (x_1, x_2) in a system of coordinates with angle ψ between the axes. In this system, an element of area is $\sin\psi\, dx_1\, dx_2$; so if we denote this element of area by da we can write the integral as

$$\frac{1}{\sin\psi}\iint \frac{\sin\omega r_{12}}{\omega r_{12}}\, da = \frac{1}{\sin\psi}\int_0^\infty \frac{\sin\omega r_{12}}{\omega r_{12}}\, 2\pi\rho\, d\rho,$$

if we take the two lines as infinitely long. But $\rho\, d\rho = r_{12}\, dr_{12}$. Hence the integral equals

$$\frac{1}{\sin\psi}\int_d^\infty \frac{\sin\omega r_{12}}{\omega r_{12}}\, 2\pi r_{12}\, dr_{12} = \frac{2\pi}{\sin\psi}\left[\frac{-\cos\omega r_{12}}{\omega^2}\right]_d^\infty = \frac{2\pi}{\sin\psi}\,\frac{\cos\omega d}{\omega^2}. \quad \ldots (36)$$

We have put $\cos\omega r_{12}=0$ at the upper limit, although it would appear to have an indeterminate value between -1 and $+1$. But we can verify that the zero value is correct if we proceed properly to the limiting case of infinitely long uniform lines of scattering matter, e.g. by starting with a Gaussian distribution of density along each line, and letting the width of this distribution tend to infinity.

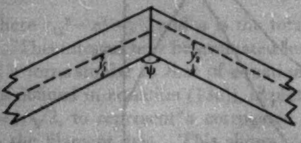


Figure 2. Strips meeting end-to-end.

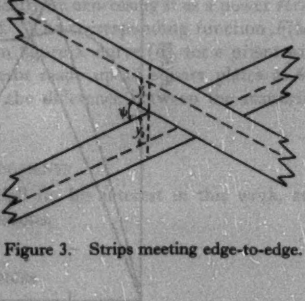


Figure 3. Strips meeting edge-to-edge.

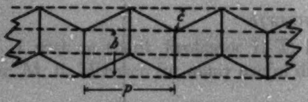


Figure 4. Corrugated filaments.

We consider first *strips meeting end-to-end* at an angle ψ. The integral (36) for the semi-infinite lines is, by symmetry, one quarter of the value for two infinite lines. For one line-element, width dy_1, at height y_1 above the lower edge of one strip (figure 2) and another line-element, width dy_2, and height y_2 above the lower edge of another strip, the contribution to the integral is

$$\frac{\pi}{2\sin\psi}\frac{\cos\omega(y_1-y_2)}{\omega^2}\,dy_1\,dy_2$$

and the whole interaction is therefore

$$\mathbf{P}_i(\omega)=\frac{\pi}{2\sin\psi}\frac{1}{a^2b^2}\int_0^b\int_0^b\frac{\cos\omega(y_1-y_2)}{\omega^2}\,dy_1\,dy_2=\frac{\pi}{2a^2\omega^2\sin\psi}\left\{\frac{\sin\frac{1}{2}b\omega}{\frac{1}{2}b\omega}\right\}^2. \qquad \ldots\ldots(37)$$

Secondly, *if the strips meet end-to-face* it is as though we had made one strip infinitely long and the other semi-infinite in the previous example, and therefore the interaction term comes out twice as large as (37).

Finally, we consider *strips meeting edge-to-edge*, as in figure 3. We consider two infinite line-elements at distance y_1 and y_2 from the edges that touch, and obtain the interaction integral as

$$\mathbf{P}_i(\omega)=\frac{2\pi}{\sin\psi}\frac{1}{a^2b^2}\int_0^b\int_0^b\frac{\cos\omega(y_1+y_2)}{\omega^2}\,dy_1\,dy_2=\frac{2\pi}{\sin\psi}\frac{\cos\omega b}{\omega^2a^2}\left\{\frac{\sin\frac{1}{2}\omega b}{\frac{1}{2}\omega b}\right\}^2. \qquad \ldots\ldots(38)$$

These interaction functions both fall to zero at $\omega b=2\pi$, and are very small for all greater values of ω; the second also has zeros at $\omega b=\pi/2$ and $3\pi/2$.

If this edge-to-edge function is added to the expression (13) for the scattering by long thin strips, the result is

$$P(\theta)=\mathbf{P}(\omega,a,b)=\frac{\pi}{ab\omega^2}\left[\int_0^{b\omega}\Lambda_1(x)\,dx+f\cos\omega b\left\{\frac{\sin\frac{1}{2}\omega b}{\frac{1}{2}\omega b}\right\}^2\right]$$

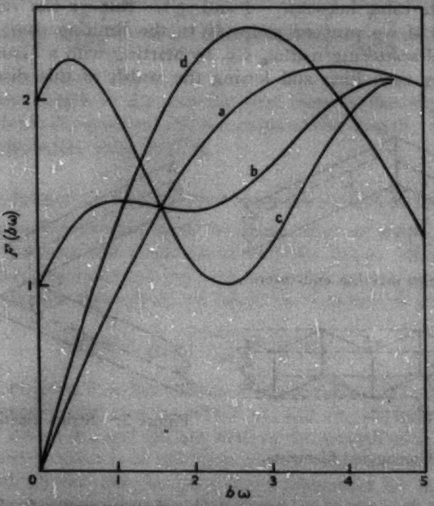


Figure 5. The function $F(b\omega)$: a, for long thin strips, width b ; b, for long strips with edge-to-edge interaction such that $f=1$; c, the same with $f=2$; d, for long prisms with $c=b/\sqrt{2}$, or the corresponding corrugated filaments.

where $f = 2b/a \sin\psi$. The part in square brackets is $F(b\omega)$ (as defined in equation (13*a*)), and is shown in figure 5, curves (b) and (c), for $f = 1$ and $f = 2$ respectively. This form of plot has been found useful by M'Ewen and Pratt (1957) for showing up the effects of interaction.

§ 9. CORRUGATED (OR ZIG-ZAG) FILAMENTS

A corrugated filament (figure 4) is a periodic structure, and we shall show that, provided the period p is less than half the wavelength of the light, the scattering is the same as that due to a prism whose cross section is a projection of the corrugated filament on to a plane perpendicular to its length.

The volume integral that occurs in equation (1) is the Fourier transform of the shape of the scattering particle. If a function of three variables (x, y, z) has periodicity p in the z direction, then the corresponding Fourier transform only has an appreciable value on the planes, in (h, k, l) space, whose equations are $l = 0$, $l = \pm 2\pi/p$, $l = \pm 4\pi/p$, etc. But $l = (4\pi/\lambda) \sin \frac{1}{2}\theta \cos\alpha$ (equations (4) and (5)): hence it cannot be greater than $4\pi/\lambda$. This means that if $p < \lambda/2$ (i.e. $2\pi/p > 4\pi/\lambda$), the only part of the Fourier transform that can have any effect on the light scattering is that for which $l = 0$.

Now this part of the Fourier transform depends solely on the projection of the particle on the (x, y) plane, and not on the distribution of matter along the z direction. It follows therefore that the corrugated filament whose period is less than half the wavelength will give the same scattering as a prism whose cross section is the same as the projection of the corrugated filament. This is shown by the broken lines in figure 4.

If this cross section is a rectangle with sides b and c, the first approximation to the scattering, for a very long filament, is given by equation (9), which becomes:

$$P(\theta) = \mathsf{P}(\omega, a, b, c) = \frac{\pi}{a\omega}\frac{1}{b^2c^2}\int_{-b}^{b}\int_{-c}^{c} J_0(\omega r_{12})(b - |x|)(c - |y|)\,dx\,dy \qquad \ldots\ldots(39)$$

where $r_{12}^2 = x^2 + y^2$, and a is the total length of the prism.

This integral may be evaluated numerically by expressing it as a power series in a form similar to that of equation (34). The corresponding function $F(b\omega)$ (as defined in equation (13*a*)), is plotted in figure 5, curve (d), for a prism with $c = b/\sqrt{2}$, to represent a corrugated filament made up of square plates at 45° to the filament axis. This shows clearly the difference between the scattering for a corrugated filament and a flat strip.

ACKNOWLEDGMENTS

I wish to thank Professor J. T. Randall for his interest in this work, and Dr. M. B. M'Ewen for many useful discussions.

REFERENCES

DEBYE, P., *Ann. Phys. Lpz.*, 1915, **46**, 809.

GUINIER, A., and FOURNET, G., 1955, *Small-angle Scattering of X-rays* (New York: Wiley).

M'EWEN, M. B., and PRATT, M. I., 1957, *Trans. Faraday Soc.*, **53**, 535.

POROD, G., 1949, *Acta Phys. Austr.*, **2**, 255.

The Interaction of Electrons with Lattice Vibrations

BY P. G. HARPER

Division of Electrotechnology, National Standards Laboratory, Commonwealth Scientific and Industrial Research Organization, Sydney, Australia

MS. received 18*th December* 1956

Abstract. Simultaneous equations are derived for the frequency and zero-point energy of normal modes of lattice motion resulting from the electron–lattice interaction.

§ 1. INTRODUCTION

IN considering the interaction of the lattice vibrations with free electrons only the harmonic terms in the lattice energy are retained because they are sufficient to describe normal modes. The interaction with electrons, however, besides modifying the harmonic motion provides also a source of anharmonicities for the new modes. It is necessary to disregard these in treating the changed phonon frequencies just as one disregards the intrinsic anharmonic terms in defining the free normal modes. Indeed, it would be inconsistent to treat the electron–lattice interaction anharmonicities without also including the latter.

A simplification follows immediately; namely it is sufficient to treat the electrons in interaction with a single normal mode (i.e. for some given **q**) for, by assumption, the resulting lattice motion, treated to harmonic order only, is independent of all other modes. It is next noted that the most general change of harmonic motion is that of a frequency shift together with a 'displacement of origin', i.e. a change in zero-point energy. In terms of the matrix representing the phonon amplitudes, the latter is represented by the addition of diagonal elements.

§ 2. THE EQUATIONS OF MOTION

A convenient description (though not the only one) is that of electron–hole pairs introduced in a previous paper (Harper 1956, to be referred to as I). The coupled equations of motion for photons and pairs are given there (equation (9))

$$\left.\begin{aligned}
&\text{(i)}\ [\rho, H] = \hbar\omega\rho + iD(b_+ - b_-^\dagger)\xi \\
&\text{(ii)}\ [\rho^\dagger, H] = -\hbar\omega\rho^\dagger + iD(b_+^\dagger - b_-)\xi \\
&\text{(iii)}\ [b_+, H] = \epsilon b_+ - iD\sum_\omega \rho \\
&\text{(iv)}\ [b_-, H] = \epsilon b_- - iD\sum_\omega \rho \\
&\text{(v)}\ [\xi, H] = -2iD\{(b_+ - b_-^\dagger)\rho + (b_+^\dagger - b_-)\rho^\dagger\}
\end{aligned}\right\} \qquad \ldots\ldots(1)$$

where

$$\rho = \sum_{\mathbf{k}} a^\dagger_{\mathbf{k}-\mathbf{q}} a_{\mathbf{k}} \delta(\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}-\mathbf{q}} - \hbar\omega)$$

$$\rho^\dagger = \sum_{\mathbf{k}} a^\dagger_{\mathbf{k}} a_{\mathbf{k}-\mathbf{q}} \delta(\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}-\mathbf{q}} - \hbar\omega)$$

are functions of **q**, ω.

The variable

$$\xi = [\rho, \rho^\dagger] \equiv \sum_{\mathbf{k}} (n_{\mathbf{k}-\mathbf{q}} - n_{\mathbf{k}}) \delta(\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}-\mathbf{q}} - \hbar\omega)$$

and b_+, b_- are the amplitudes of normal modes $\mathbf{q}$, $-\mathbf{q}$ respectively, of energy ϵ: H is the total Hamiltonian and D the interaction parameter as defined by Fröhlich (1952).

It is convenient to abbreviate $b_+ - b_-^\dagger$, $b_+^\dagger - b_-$ to B, $B^\dagger$ respectively so that $[B, B^\dagger]=0$. The assumption of harmonic motion for the lattice now implies that in the representation which diagonalizes H, B need be considered to have matrix elements only between n, n and n, $n-1$ where n is the new phonon excitation number. The diagonal element will be denoted by b while the non-diagonal element $B_{n,n-1}$, say, will be abbreviated simply to B. Similarly, the operator $B^\dagger$ will have corresponding elements $\bar{b}$ and $B^\dagger$. It will presently appear that $\bar{b}=-b$, i.e. b is imaginary. It can then be shown that $B^\dagger=-\bar{B}$, i.e. B is also imaginary.

Consider equations (1) in the H-representation: denote $E_n - E_{n-1}$ by E (new phonon energy), and let $\rho_{n,n-1}$, $\rho^\dagger{}_{n,n-1}$ be abbreviated to ρ, $\rho^\dagger$ respectively, their common (real) diagonal element being r. Let the diagonal element of the operator ξ be ξ and the non-diagonal element $\xi_{n,n-1}$ be x. It was remarked in I that neglect of x in the equation (11) of I was related to neglect of correlation of electron particle states. Then (i), (ii) and (v) taken between adjacent phonon states become

$$\text{(i)}\quad (E-\hbar\omega)\rho = iD(bx+B\xi)$$

$$\text{(ii)}\quad (E+\hbar\omega)\rho^\dagger = iD(\bar{b}x+\bar{B}\xi)$$

$$\text{(v)}\quad Ex = -2iD(\bar{b}\rho+b\rho^\dagger).$$

Substituting from (i) and (ii) into (v)

$$x = 2D^2\left\{\frac{2|b|^2}{E^2-(\hbar\omega)^2}x + \frac{2\bar{b}B}{E^2-(\hbar\omega)^2}\xi\right\}$$

or

$$x = \frac{4D^2 b\bar{B}}{E^2-(\hbar\omega)^2-4D^2|b|^2}\xi \qquad \ldots\ldots(2)$$

and therefore

$$\rho = \frac{iDB(E+\hbar\omega)}{E^2-(\hbar\omega)^2-4D^2|b|^2}\xi. \qquad \ldots\ldots(3)$$

From (iii) and (iv)

$$B = \frac{2iD\epsilon}{\epsilon^2-E^2}\sum_{\omega}\rho. \qquad \ldots\ldots(4)$$

Substituting (3) into (4)

$$\left\{\epsilon^2-E^2-2D^2\epsilon\sum_{\omega}\frac{(E+\hbar\omega)\xi}{(\hbar\omega)^2-E^2+4D^2|b|^2}\right\}B=0. \qquad \ldots\ldots(5)$$

Since B does not vanish (except for the limiting case of infinitely heavy ions) there results an equation for E in terms of the known function ξ, in this case supposed due to the Fermi ground state of the electrons, and the so far unknown b which is related to the zero-point energy U of the phonons by $U=-|b|^2E$. Of course, ξ, E, D, b and ϵ all depend upon $\mathbf{q}$, the wave vector of the excited mode.

A second equation is obtained by taking diagonal elements of (1), thus

$$r = -\frac{iD}{\hbar\omega}(b\xi+Bx), \qquad \epsilon b = 2iD\sum_{\omega} r$$

which, using (2) and (3), give,

$$\left[\epsilon-2D^2\sum_{\omega}\frac{\xi}{\hbar\omega}\left\{1-\frac{4D^2|B|^2}{(\hbar\omega)^2-E^2+4D^2|b|^2}\right\}\right]b=0. \qquad \ldots\ldots(6)$$

If b does not vanish, (5) and (6) constitute a pair of simultaneous equations for the energies E and $2D|b|$.

§ 3. DISCUSSION

In I, the non-diagonal matrix elements x of the Fermi factor $n_{\mathbf{k}-\mathbf{q}}-n_{\mathbf{k}}$ were disregarded; this is seen to be a genuine solution for the case $b=0$, equation (5) reducing to

$$\epsilon^2-E^2=2D^2\epsilon\sum_{\omega}\frac{\xi}{\hbar\omega-E},$$

which was discussed in I. It was noted there that the vanishing of the resonance denominator $\hbar\omega-E$ described true electron–phonon scattering.

There is now seen to be a second solution $b\neq0$, whose ground-state energy is lower than that of the scattering state by $|b|^2E$ and for which $x\neq0$. This state no longer permits scattering since $\hbar\omega-E$ is replaced by $(\hbar\omega)^2-E^2+4D^2|b|^2$, an expression similar to that describing damping in radiation theory. The situation is analogous to the well-known case of a static sinusoidal potential imposed on free electrons, where it was shown in I that exact treatment of x removed the resonance denominator and replaced the energy divergence by a finite discontinuity. Indeed, in the limit of infinitely heavy ions, $\epsilon\to0$, $B\to0$ and equations (i), (ii), (v) lead to

$$E_w^2=(\hbar\omega)^2+4D^2|b|^2.$$

This is equation (8) of I, where the energy gap is given explicitly by $2|a|=2D|b|$.

In the limit of free phonons, $D\to0$, $b\to0$ and from (5) $E\to\epsilon$ as it should.

It has been tacitly supposed that phonons were excited from the ground state and in that case $\xi(\mathbf{q},\omega)$ is calculated from the free electron distribution, either Fermi or Fröhlich, as in I. For higher excited states, ξ_n is determined in terms of $B_{n,n-1}$ by the equation $[\rho,\rho^\dagger]=\xi$, which substituting for ρ, $\rho^\dagger$, gives

$$|B_{n,n-1}|^2\xi_n^2-|B_{n+1,n}|^2\xi_{n+1}^2=\frac{\xi_n}{4D^2E\hbar\omega}[(\hbar\omega)^2-E^2-4D^2|b|^2]^2.$$

The values for $B_{n,n+1}$ are derived from those normally used in free phonon excitation, since, by definition, they refer to non-interacting phonons.

The electronic motion cannot, of course, be described independently of that of the phonons since it co-operates to produce it in much the same way as, for example, is supposed in the adiabatic approximation to the motion of the lattice. The correct number of degrees of freedom is preserved since the electronic states can be enumerated (though not fully described because of the correlation mentioned) in the usual way by occupation numbers $n_{\mathbf{k}}$, and the formal description of the lattice motion, namely, plane waves, is unchanged.

Equations (5) and (6) remain to be examined for (i) reality of E, i.e. stability of the lattice, and (ii) magnitude of $|b|$, i.e. reduction of ground-state energy. These questions will be dealt with in a future publication.

REFERENCES

FRÖHLICH, H., 1952, *Proc. Roy. Soc.* A, **215**, 291.
HARPER, P. G., 1956, *Proc. Phys. Soc.* B, **69**, 1095.

Metal Fatigue at High Frequency

By E. A. NEPPIRAS

Mullard Research Laboratories, Salfords, Redhill, Surrey

Communicated by R. W. B. Stephens; MS. received 16*th May* 1956,
and in revised form 23*rd September* 1956

---

§ 1. Introduction

In recent years the problem of the failure of metal structures when subjected to repeated fluctuating stresses has assumed considerable importance and much new work is now going on in an attempt to obtain a better understanding of the true nature of fatigue damage and to improve testing techniques.

In both research and routine testing it is generally necessary to apply stress alternations for up to 10^8 reversals. Conventional machines for investigating fatigue damage operate at repetition frequencies between 25 and 100 c/s and a single test run may therefore occupy several weeks. Many attempts have been made to devise a technique which is both rapid and reliable, but although some have shown promise none of the accelerated testers produced so far can be regarded as thoroughly reliable.

Several workers have devised resonant electromagnetic vibrating systems capable of producing stress alternations at frequencies up to a few kc/s. Measurements have also been reported in which compression–tension stresses were obtained by means of resonant pneumatic vibrators operated at a range of frequencies up to about 10 kc/s (Jenkin and Lehmann 1929). These measurements appear to show that for most metals the fatigue limit increases with increasing frequency to a peak at about 8 kc/s, decreasing again at still higher frequencies. The apparatus employed could not easily be adapted to industrial testing. Also the interpretation of the measurements involves uncertain assumptions which render the results very much open to criticism.

The possibility of using a simplified high-frequency vibrator, based on resonant ultrasonic transducers, has also been noted in the literature (Mason 1950, Neppiras 1953 a, b). Such vibrators are capable of producing large alternating stresses at frequencies up to several megacycles per second when suitable transducers and focusing devices are used. The most useful and convenient frequencies for fatigue testing appear to lie in the high sonic and low ultrasonic ranges (about 5 to 50 kc/s) and as generators in this range magnetostrictors are decidedly more valuable than the other available types of transducer. They are more robust, less temperature sensitive and capable of withstanding greater strains.

In the work to be described, metal samples were subjected to tension–compression stresses above the fatigue limit at frequencies around 18 kc/s produced by a resonant magnetostriction transducer. The work has shown that the high-frequency technique possesses a number of advantages over conventional methods of fatigue analysis. But the measurements confirm that at this frequency the fatigue limit is appreciably higher than that obtained from low-frequency measurements. This frequency effect would complicate the design of a practical accelerated testing machine.

§ 2. THE VIBRATOR

The apparatus used in the tests has the advantage of being mechanically very simple—simpler even than existing low-frequency machines. It consisted of a resonant magnetostriction transducer of conventional design coupled to the fatigue sample through a mechanical transmission line using resonant tapered elements designed to transform up the alternating stresses from the transducer to a value above the fatigue limit in the specimen. The transducer and coupling elements operated at stress levels well below their fatigue limits, but the sample could be subjected to stresses capable of breaking it in a very short time.

The essentials of this system are shown in figure 1. The transducer was of the laminated dumbell type and the coupling stubs were each halfwave resonant elements. The test pieces were cut to be also approximately resonant at the transducer frequency and were given a symmetrically tapering section, the diameter at the centre ($\frac{3}{16}$ inch) being about half that at the ends. The sample was hard-soldered to the free end of the final transformer in the mechanical line and the whole assembly was supported on a single metal flange fixed at the displacement node of the first transformer, this support forming one wall of a water-cooling enclosure for the transducer. The separate sections of the transmission line were tapered exponentially and the resonant dimensions calculated in accordance with simple theory (Neppiras 1953 c). Under test conditions the mechanical Q of the system was sufficiently high to ensure that the stress reversals were accurately sinusoidal. The single flange mounting system adopted ensured that the tests were done under zero static load conditions. There is no possibility of any direct axial loading developing during the tests. The transducer was driven from a variable frequency oscillator-amplifier; adjustment of the drive power enabled the stress level in the sample to be set to the desired value.

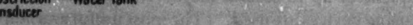


Figure 1. Experimental arrangement for investigating metal fatigue at high frequency.

Figure 2 shows a plot of the distribution of the oscillatory amplitude of vibration along the transmission line, the amplitude being measured with a microscope. The derived stress amplitudes are also plotted; these are proportional to the slopes of the amplitude–distance curves. The stress in the system rises to a high peak near the centre of the sample, where the fatigue break eventually occurs. A measurable stress was obtained at the free end of the sample. This is accounted for by loading due to spray from the cooling water as it emerged.

High-frequency straining of the sample at high level is accompanied by considerable power absorption due to mechanical hysteresis, so that for controlled measurements efficient cooling is necessary. Preliminary measurements showed

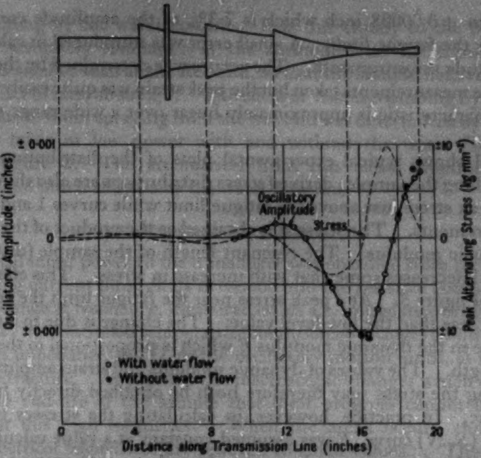


Figure 2. Distribution of particle motion and stress in the mechanical transmission system.

that neither an air blast nor a liquid spray applied to the outer surface was entirely adequate; the specimen was therefore bored with a small ($\frac{1}{16}$ inch diameter) hole which was made to carry a fast continuous flow of water. The axial hole was also taken through the final coupling stub as far as its node, as this provided a convenient place to introduce the feed pipe to avoid excessive damping (see figure 1). This method of cooling was found to be very efficient; the water flow rate could be adjusted to keep the outer surface of the sample within a few degrees of the water temperature even at the highest vibration intensities. As most of the heat dissipated was carried away by the water flow, temperature and flow rate measurements also provided a convenient and quite accurate measure of the power absorbed by the sample.

§ 3. MEASUREMENT OF STRESS

The fatigue tests were carried out on samples of brass, composition Cu 67%; Zn 33%; Pb 0·3%; Sb, As 0·02%; Sn 0·01%; Fe, Ni 0·005%, cold drawn with the surface fine machined. The main observations consisted in measuring the time required for fracture at a number of values of peak stress near and above the fatigue limit. From these measurements the (S, N) curve (the curve relating the peak applied stress with the number of reversals to fracture) was plotted and the fatigue limit estimated.

The distribution of strain in the sample was deduced from microscope measurements of the peak-to-peak oscillatory amplitude ξ taken at a number of points along its length x. In all measurements of amplitude it was the longitudinal component which was obtained. The peak axial strain was then obtainable directly as half the slope of the (ξ, x) curve. This method of measurement had the advantage that both the position and magnitude of the peak strain were obtained directly and this rendered the measuring technique independent of the precise shape and size of the sample. The microscope used in these measurements recorded the oscillatory amplitude to ±0·00008 inch which is 2·3% of the amplitude corresponding to a stress near the fatigue limit. A small error was introduced in calculating strain from amplitude measurements. The accuracy is determined by the total number of amplitude measurements taken but the peak strain was quite easily obtained since the (ξ, x) characteristic is approximately linear over a wide range near the strain antinode.

Figure 3 shows typical experimental plots of the distribution of oscillatory amplitude along the sample; derived stress distributions are also shown. Curves 3 refer to a peak stress just above the fatigue limit while curves 1 and 2 refer to low level measurements. The stress S is obtained as the product of the strain and the dynamic elastic modulus. The resonant length of the sample (distance between stress nodes) decreases somewhat with increase in stress. This effect is shown in the curves of figure 3. At a peak stress near the fatigue limit the resonant length is about 6% less than the low-level value. The change is due to a corresponding small change in the dynamic modulus E which is proportional to the square of the resonant length. The values of dynamic modulus and strain amplitude to be used in calculating the stress may therefore both be obtained directly from the (ξ, x) characteristic. In practice, however, in calculating the stresses to be used for plotting the (S, N) curve, E was assumed constant at a value calculated from the resonant length of an untapered sample vibrated at a peak stress near the fatigue limit. This assumption is justified considering the narrow range of stress involved and the fact that the changes are still quite small in this region.

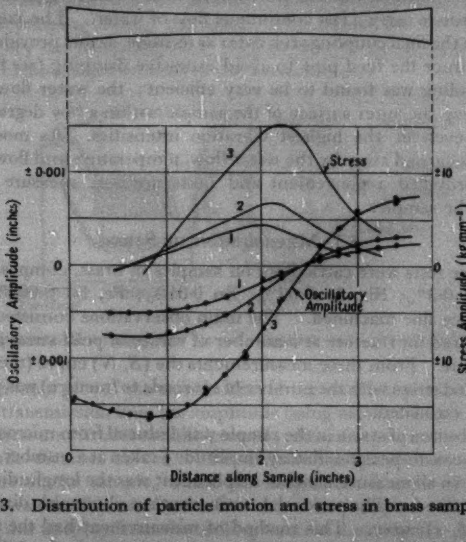


Figure 3. Distribution of particle motion and stress in brass sample piece.

The particular form of the (ξ, x) and (S, x) curves requires some explanation. The considerable reduction in oscillatory amplitude along the sample is only partially accounted for by attenuation due to internal damping in the material and to viscous damping by the water coolant. It is due chiefly to the fact that the sample is not accurately resonant; at the antinode near the transformer the sectional area is smaller than at the free end, so that the sample behaves as a step-down transformer of ratio about 2/3. The effect of viscous damping by the coolant and damping due to water spraying from the free end was determined by comparing measurements taken at low power with and without the water flow. These are shown in figure 2. The attenuation from this cause amounts to about 1 dB and this is not much affected by the water flow rate.

§ 4. EXPERIMENTAL RESULTS

The electrical input power to the transducer was obtained from measurements of the alternating current through its drive coil and the resistive component of its electrical impedance at resonance, measured on an a.c. impedance bridge. It was found that under the controlled temperature condition no changes could be detected in the electrical impedance, the Q, or the oscillatory motion, until, at most, a few seconds before fracture. Apparently the macroscopic changes which occur are so confined to the end-life of the sample that their progress cannot be recorded conveniently at this high frequency. When the break occurs, the oscillatory motion and electrical impedance are reduced to small values. The impedance drops almost to the clamped electrical impedance of the transducer, since the fatigue crack occurs near a displacement node and the sample then becomes a quarter-wave stub which serves as a very effective mechanical clamp at the transducer resonance. This impedance change might therefore be used as a criterion of fatigue. In all the samples tested fracture took the form of a lateral crack which occurred very near the point of maximum strain, the stress amplitude at the point of fracture being always within $-1{\cdot}5\%$ of the maximum stress.

The experimental plots and the (S, N) curve for the brass samples used in these tests are given in figure 4. The scatter in the experimental points is about that to be expected, bearing in mind the probable error in fixing S which is around ± 2–3%. At points corresponding to high values of stress an additional inaccuracy is introduced by the extremely short time available for recording the observations. The uncertainty in recording N was about $\pm 10^5$ cycles† and this is, of course, independent of N. For this reason no observations were recorded for S greater than about 17 kg mm^{-2} where fatigue occurs in about 2×10^6 reversals

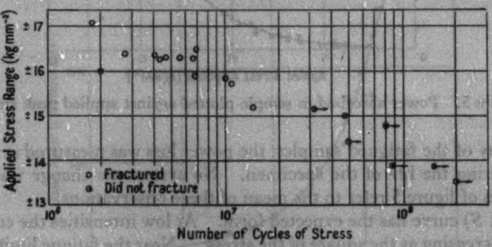


Figure 4. The (S, N) curve.

---

† This corresponds to an estimated maximum uncertainty of ± 5 seconds in the timing of a run.

(i.e. about two minutes). At points near the fatigue limit, on the other hand, the uncertainty in the time measurement is much less serious than the error in fixing S.

The (S, N) curve gives the fatigue limit, on the basis of 10^8 reversals, as $\pm 15\,\mathrm{kg\,mm^{-2}}$ within about $\pm 0{\cdot}5\,\mathrm{kg\,mm^{-2}}$. This is substantially higher than the mean of published figures for similar material in tests carried out at low frequencies on the Haigh tester ($\pm 13{\cdot}1\,\mathrm{kg\,mm^{-2}}$ with scatter of $\pm 1{\cdot}5\,\mathrm{kg\,mm^{-2}}$). This frequency dependence of the fatigue limit follows the trend observed by other workers. It means that an accelerated tester would need to be calibrated relative to the low frequencies appropriate to the actual duty cycle of the component under test.

§ 5. ENERGY ABSORPTION

An alternative method of fixing the fatigue limit is suggested by calorimetric measurements of mechanical loss in the sample using the cooling water system as a constant flow calorimeter. The rate of energy loss was calculated from measurement of the flow rate and temperature rise of the cooling water. This was done for each fatigued sample and over the full stress range below the fatigue limit, thus providing data for a plot of power dissipated W as a function of the peak stress S (see figure 5). The water flow rate was increased to keep the surface temperature down to within a few degrees of the ambient temperature; the heat lost to the atmosphere from the outer surface could then be neglected.

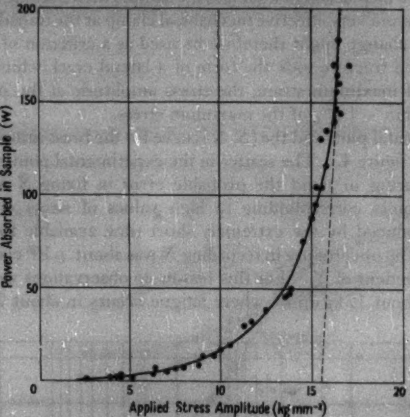


Figure 5. Power absorbed in sample plotted against applied peak stress.

For many of the fatigued samples, the power loss was measured at one minute intervals during the life of the specimen. No systematic change was observed, and the plots of figure 5 refer to the mean of these observations.

The (W, S) curve has the expected form. At low intensities the energy losses are small, increasing as the square of the stress. Near the fatigue limit the energy absorbed in the sample increases rapidly corresponding to a significant enlargement of the mechanical hysteresis loop (i.e. in this region the imaginary part of the complex dynamic modulus is a rapidly increasing function of the stress). Above the fatigue limit, the (W, S) curve tends to linearity and intersects the stress axis at $\pm 15{\cdot}5\ \mathrm{kg\ mm^{-2}}$. This approximates to the fatigue strength at 10^8 reversals, after which we can assume that the (S, N) curve has become horizontal or nearly so. Power absorption measurements may thus provide an alternative means of fixing the fatigue limit.

Strictly, the figures for power dissipation require to be corrected for the fact that since the cooling water is introduced in the nodal plane of the transformer the measured dissipation includes a contribution from the heat losses in part of the transformer stub. Also, as noted above, some energy is transferred to the liquid as viscous loss. Both corrections must be small and have not been applied. They could not, in any case, affect the value obtained for the fatigue limit since the correction terms are slowly increasing functions of stress near the fatigue limit, as compared with the very rapid changes shown in the (W, S) plots.

This technique of using energy absorption measurements to fix the fatigue limit may prove a valuable one. At high frequency the losses at stresses near the fatigue limit are great enough to be easily measured to the accuracy required for a fatigue determination. The loss can be obtained either calorimetrically or from electrical impedance measurements on the transducer, but the calorimetric measurement is more direct, more rapid, and on the whole less liable to error. With the samples used in the present work, a reliable measurement of loss could be obtained within a few seconds of setting the stress. Since understressing for short periods appears to have little effect on the fatigue limit, it seems possible that a succession of loss measurements on a single sample might provide sufficient data for a reliable estimate of the fatigue limit. The stress would be increased in small steps to a value above the fatigue limit, which could then be deduced from the curve relating power loss and stress. The value of this method of fatigue measurement has not yet been established but if it proves practicable it will enable a complete fatigue measurement to be carried through in only a few minutes.

In figure 6 the total electrical input power to the transducer is shown as a function of the stress level in the sample. As to be expected, the relation is a square law only at low intensities. At stresses near and beyond the fatigue limit the power supplied increases rapidly. Approximately 1 kilowatt input power is required to fatigue a sample under the conditions of the tests. The curves of

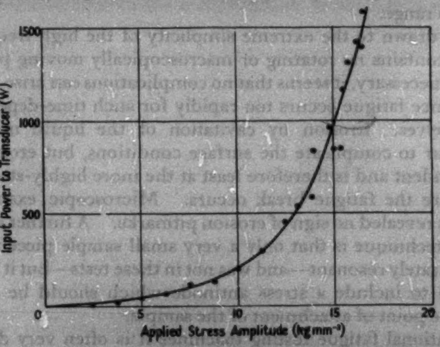


Figure 6. Electrical input power plotted against applied peak stress.

figures 5 and 6 show that the electromechanical efficiency of the system (defined as (power absorbed in specimen/input electrical power)) is constant at low levels but increases as the stress is taken up towards the fatigue limit. But even above the fatigue limit the efficiency is only 9–12%. Under these conditions any small changes in the specimen damping during the test run would have little effect on the magnitude of the stress established in it.

These observations of the rate of energy loss in fatigued samples emphasize the need for efficient cooling if the measurements are to be directly related to low frequency measurements. If cooling is omitted, very high temperatures are produced in the highly stressed regions, increasing as the breaking point is approached. The temperature near the stress antinode may reach red, or even white, heat in a few seconds in a lossy material. Under these conditions, the breaking point is approached comparatively slowly, since the attenuation accompanying the increased losses acts as a brake on the progress of fatigue. The increased losses result in a greatly reduced Q and at constant input power the stress in the samples is correspondingly reduced. With a sample in the form of a rod of small diameter/length ratio (so that heat transfer by conduction axially is small in comparison with radiation from the surface at high temperature) the surface temperature gives an approximate measure of the rate of generation of heat by internal loss at any point. It is found that the zone of very high temperature is confined to a very restricted region near the stress antinode. The temperature variation along the rod is very different from the stress-squared distribution which would be obtained at lower stresses. These observations afford striking confirmation that the energy absorbed in the sample—the area enclosed by the hysteresis loop—increases rapidly as a function of both the stress amplitude and time as the fatigue limit is approached.

§ 6. ADVANTAGES AND LIMITATIONS OF THE HIGH FREQUENCY TECHNIQUE

The chief advantage of high frequency testing is, of course, its rapidity: a normal fatigue run of 10^8 reversals can be carried through in about 1½ hours on the magnetostriction tester compared with several weeks when conventional machines are used. The limit of laboratory experience of fatigue testing is at present about 10^9 reversals. The high frequency machine will obviously allow an enormous extension of this range.

Attention is drawn to the extreme simplicity of the high frequency vibrator. The apparatus contains no rotating or macroscopically moving parts. Although liquid cooling is necessary, it seems that no complications can arise due to corrosion of the sample since fatigue occurs too rapidly for such time-dependent effects to manifest themselves. Erosion by cavitation of the liquid might have been expected to occur to complicate the surface conditions, but erosion is primarily amplitude-dependent and is therefore least at the more highly-stressed regions of the sample where the fatigue break occurs. Microscopic examination of the fatigued samples revealed no sign of erosion pitmarks. A further advantage of the high frequency technique is that only a very small sample piece is required. It need not be accurately resonant—and was not in these tests—but it must, of course, be long enough to include a stress antinode which should be at a reasonable distance from the point of attachment of the sample.

With conventional fatigue testing machines it is often very difficult to make tests under repeated tension–compression stresses without introducing appreciable stresses due to flexure. With the magnetostriction tester, provided reasonable care is taken in aligning the sample axially with respect to its associated transformer, the vibrations obtained are always accurately longitudinal and free from flexure. A sound, rigid, bond between the sample and its transformer is, however, quite essential. Easily detachable joints, such as screw threads and tapers, were found to be unsatisfactory at this junction, and in the measurements recorded here the samples were hard-soldered to the transformer using a localized flame heating. For some metals it would obviously not be convenient to use soldered or brazed joints; in such cases the sample could without difficulty be formed integral with the transformer, provided sufficient material was available.

The present work has confirmed that the fatigue limit is a function of the operating frequency. This would complicate the design of an accelerated tester, which would need to be calibrated relative to the low frequencies appropriate to the actual duty cycle of the component under test.

ACKNOWLEDGMENTS

The work reported was carried out at the Mullard Research Laboratories and constitutes the preliminary stage of a proposed long term investigation of metal fatigue at high frequency.

The author wishes to acknowledge helpful advice from members of the Laboratories. Thanks are also due to the Directors of the Mullard Radio Valve Company for permission to publish this paper.

REFERENCES

JENKIN, C. F., and LEHMANN, G. D., 1929, *Proc. Roy. Soc.* A, **125**, 83.

MASON, W. P., 1950, *Piezoelectric Crystals and their Application in Ultrasonics* (New York: van Nostrand), p. 161.

NEPPIRAS, E. A., 1953 a, *Research*, **6**, 271; 1953 b, *Metal Treatm.*, **20**, 391; 1953 c, *J. Sci. Instrum.*, **30**, 72.

Relaxation Processes and Inertial Effects
I: Free Rotation about a Fixed Axis

By R. A. SACK†
Department of Theoretical Physics, University of Liverpool
and
British Rayon Research Association, Wythenshawe, Manchester

MS. received 20th November 1956

Abstract. The limitations of the usual concepts of relaxation processes for very short periods are discussed, and it is shown that in the limit of very high frequencies the product of the loss factor and frequency must tend to zero. For a consistent kinetic treatment the probability must be considered in phase space, not merely in configuration space; for Cartesian coordinates the stochastic (generalized Liouville) equations can be simplified by a Fourier transformation with regard to the velocities.

By means of the simplified equations the dielectric properties are determined for a system of rigid dipoles rotating about fixed axes in a viscous medium. The results are equivalent to, but more compact than, those derived for the same problem by E. P. Gross; for strong viscous damping they agree with Debye's relaxation formulae, except at very high frequencies. The polarization can be described by an infinite discrete set of relaxation times with amplitudes which alternate in sign; the longest relaxation time is identical with that calculated by Debye. The logarithm of the function expressing the dielectric after-effect follows a law similar to the Ornstein–Fürth generalization of Einstein's law of diffusion. The analytical solutions obtained by Gross for abrupt collisions between dipoles and heat-bath are also re-derived by a somewhat easier method.

§ 1. Introduction

In the theory of relaxation phenomena it is usually assumed that if a physical system is subjected to an external *field* which is not so large as to cause saturation effects, the contribution to the induced response (denoted in the following by the term ***moment***) due to each elementary mechanism tends exponentially towards its equilibrium value. Thus, if only one type of mechanism is operative, the moment $P(t)$ of a system to which the constant field F_0 is applied at time $t=0$ will be

$$P(t)=P_0(1-e^{-t/\tau})=k_0F_0(1-e^{-t/\tau}), \quad (t>0) \qquad \ldots\ldots(1.1)$$

where

$$P_0=k_0F_0 \qquad \ldots\ldots(1.2)$$

is the equilibrium value of the moment and τ is a constant, usually known as the relaxation time. The value of the equilibrium coefficient k_0 can usually be derived from statistical mechanics, whereas τ can only be deduced from kinetic considerations. If such a system is subjected to a sinusoidally alternating field

$$F=F_0\cos\omega t=F_0\mathscr{R}(e^{i\omega t}), \qquad \ldots\ldots(1.3)$$

† Now at Department of Theoretical Chemistry, University of Cambridge.

the steady state response can be described by a complex response coefficient depending on frequency

$$k^*(\omega) = k'(\omega) - ik''(\omega) \qquad \ldots\ldots(1.4)$$

such that

$$P = F_0\,[k'(\omega)\cos\omega t + k''(\omega)\sin\omega t] = F_0\,\mathscr{R}\,(k^*(\omega)\,\mathrm{e}^{i\omega t}). \qquad \ldots\ldots(1.5)$$

The coefficient $k''(\omega)$, the loss factor, is a measure of the energy loss *per cycle* for a constant amplitude F_0; the relevant quantity for the energy loss *per unit time* is the product $\omega k''(\omega)$. For a system obeying (1.1) and (1.2) the frequency dependence of k^* is given by

$$k^*(\omega) = k_0/(1 + i\omega\tau)\,; \quad k'(\omega) = k_0/(1 + \omega^2\tau^2)\,; \quad k''(\omega) = k_0\omega\tau/(1 + \omega^2\tau^2). \qquad \ldots\ldots(1.6)$$

The loss factor $k''(\omega)$ has a maximum for $\omega\tau = 1$. Experimental curves of k'' against $\log\omega$ usually show an absorption peak broader than predicted by (1.6); this can be explained by a distribution $k(\log\tau)$ of relaxation times, each component giving a contribution according to (1.1) and (1.6), so that the approach of the moment to its equilibrium value becomes

$$P(t) = F_0 \int k(\log\tau)\,(1 - \mathrm{e}^{-t/\tau})\,d(\log\tau) \qquad \ldots\ldots(1.7)$$

instead of (1.1), and the frequency dependent response coefficient

$$k^*(\omega) = \int \frac{k(\log\tau)\,d(\log\tau)}{1 + i\omega\tau}. \qquad \ldots\ldots(1.8)$$

The distribution function $k(\log\tau)$ is generally assumed to be non-negative by nature of the assumptions made; it may contain infinities (of the Dirac δ-function type) corresponding to a set of discrete relaxation times.

It has been pointed out repeatedly that, whatever the mechanism responsible for the relaxation, the formulae (1.1) and (1.6) must break down at very short periods. Among other absurd conclusions, they predict that in a dielectric a relaxation absorption in the audio- or radio-frequency range would entail a finite absorption coefficient for visible light with complete opacity in some extreme cases such as water (Debye 1934) and an infinite rate of energy loss for an electron travelling through the material (Fröhlich and Platzman 1953). Furthermore, equation (1.1) would imply that an abrupt change of the field F would produce an instantaneous finite alteration of the rate of change dP/dt of the moment, which is impossible in view of the finite masses of the constituent particles (cf. Sack 1953 a). In general the product $\omega k''(\omega)$ must tend to zero in the limit of very high frequencies which means that in the frequency range in which $k^*(\omega)$ can be expanded in negative powers of $i\omega$, i.e.

$$k^*(\omega) = a_0 + \frac{a_1}{i\omega} + \frac{a_2}{(i\omega)^2} + \ldots, \qquad \ldots\ldots(1.9)$$

the coefficient a_1 must vanish. This is compatible with a distribution of relaxation times of the type (1.7)–(1.8) only if the function $k(\log\omega)$ can take negative as well as positive values (B. Gross 1953, Sack 1953 b), for we must have

$$a_1 = \int k(\log\tau)\,d(\log\tau)/\tau = 0. \qquad \ldots\ldots(1.10)$$

The above considerations show that in all relaxation processes inertial effects become important at sufficiently high frequencies. In the case of mechanical relaxation, where the field F and the moment P represent the stress and the strain respectively, the influence of the masses of the individual particles can usually be summed over small elements of volume and be expressed in terms of a macroscopic density. In the case of dielectric relaxation, however, where the formulae (1.1)–(1.8) express the results of the well-known Debye theory (Debye 1945), the motions of neighbouring dipoles are largely uncorrelated; the relevant inertia terms refer essentially to the individual polar molecules, and in consequence are usually neglected. It is thus of great importance to know how such inertial effects affect the dielectric properties of the material at very high frequencies.

For classical fluid dielectrics the Debye equations are derived in principle on the basis of the Smoluchowski equation which describes the change in time of a distribution function in configuration space (in this case the space of the possible directions of the rotating dipoles); in each position the polar molecule is assumed to have that mean (angular) velocity which corresponds to dynamic equilibrium between the orienting forces of the local field and the viscous drag due to the surrounding medium. In this approximation inertia effects are neglected altogether; attempts to account for them by an additional term in the original equations of motion (Rocard 1933, Dmitriev and Gurevich 1946, Powles 1948) are only partially correct. Inertia effects can be consistently treated only on the basis of the generalized Liouville equation which considers distributions in phase space where both positions and momenta (or velocities) are taken as independent variables. Whereas this procedure is standard in the theory of resonance absorption (Fröhlich 1946, Gross 1955 a), it has been applied to relaxation processes only in a few articles (cf. Harper and O'Dwyer 1955, Saito and Kato 1956). The most detailed treatment is due to E. P. Gross (1955 b) who considers the behaviour of rigid dipoles rotating about fixed axes in a viscous medium in connection with the dielectric properties of compressed polar gases. Some of Gross's results had already been given in a somewhat more compact form by the writer (Sack 1953 b), but without the derivation which was based on a Fourier transform of the generalized Liouville equation. The aim of this paper is to present in detail this alternative treatment; the basic stochastic equations and their transforms are discussed in §2; the analytic solutions for a system of dipoles rotating in a plane under several possible types of randomizing influences are derived in §3, and the results are discussed and interpreted in physical terms in §4. Consideration of models with free rotation in space or involving restoring forces will be left to subsequent publications (Sack 1957).

§2. THE KRAMERS EQUATION AND ITS FOURIER TRANSFORM

The statistical behaviour of a classical macroscopic assembly under the action of external and internal forces is described by the Liouville equation

$$\frac{DW}{Dt} \equiv \frac{\partial W}{\partial t} + \sum \left\{ \frac{\partial H}{\partial p_k}\frac{\partial W}{\partial q_k} - \frac{\partial H}{\partial q_k}\frac{\partial W}{\partial p_k} \right\} = 0, \qquad \ldots\ldots(2.1)$$

where W is a probability distribution in the phase space of $2N$ dimensions, N being the number of degrees of freedom of the system, and H is the Hamiltonian. As an equation with a number of variables of the order of 10^{20} is not tractable, it is usually necessary to modify Liouville's equation by both reducing and generalizing it; reducing, by limiting the degrees of freedom to a small, but representative set with a well-defined potential, and generalizing, by the addition of terms on the right-hand side of (2.1) to account for the mean interaction between this set and the remaining degrees of freedom (the background or heat-bath).

In the case of statistical equilibrium DW/Dt will vanish for the generalized Liouville equation, since this equilibrium is independent of the nature of the interaction between the set considered and the background. Of the various possible types of interactions, the ones mostly investigated belong to the class of *adiabatic* collisions which are of such short duration compared to other explicitly considered time intervals, such as periods of oscillation or rotation, that they can be treated as instantaneous. The present treatment will be confined to such collisions, and it will further be assumed that the background is always in thermal equilibrium. A number of possible types of collisions have been discussed by Kramers (1940), by Keilson and Storer (1952) and by Gross (1955 b); of these the most important are:

(A) *Small Collisions* (*Brownian Motion*), in which the positions of the particles are unchanged and their velocities altered by such small amounts that they can be treated as infinitesimal; correlation between successive impulses is negligible. Under the simplest assumptions, the effect of such collisions appears as a viscous retardation

$$dv/dt = -\beta v, \qquad \ldots\ldots(2.2)$$

superimposed on which are random fluctuations.

(B) *Large Collisions*, in which the position of the particles remains unaltered on impact, but all coherence between their initial and final velocities is destroyed; after each collision a particle may have any velocity with a probability according to Boltzmann's law for the given temperature.

(C) *Very Large Collisions*, which destroy coherence of position as well as of momentum; after each impact a particle may be found anywhere in phase space according to the Boltzmann distribution corresponding to the momentary Hamiltonian.

The physical interpretation of these limiting, and some intermediary, cases has been given by Gross (1955 b); it may suffice here to summarize that assumption (A) corresponds to the case of the explicitly considered particles being surrounded by a heat-bath formed by much smaller particles; in case (B) the molecules of the heat-bath are considerably heavier, whereas assumption (C) is, strictly speaking, contrary to the laws of impact, but can sometimes be taken to be a useful approximation in the case of disproportionately heavier particles forming the background.

The collision rates can, in principle, be functions of the positions and velocities, but will be treated here as constants of the system, as is usually done.

The following treatment is given in detail only for a motion along a straight line, but generalization to Cartesian coordinates in several dimensions is straightforward, provided only the rate constant β and the effective mass m are either scalar constants or else tensors with the same principal axes. Considerable complications arise, however, for curvilinear coordinates. To avoid any erroneous generalizations, we shall consider a distribution w in the space of Cartesian position coordinates x and velocities v rather than the distribution W in phase space of equation (2.1).

For movement in one dimension under the influence of Brownian collisions (case A), the stochastic equation for $w(x, v, t)$ is Kramers' (1940) generalization of the Fokker–Planck equation, namely

$$\frac{Dw}{Dt} \equiv \frac{\partial w}{\partial t} + v\frac{\partial w}{\partial x} - \frac{1}{m}\frac{\partial V}{\partial x}\frac{\partial w}{\partial v} = \beta\frac{\partial}{\partial v}\left(vw + \frac{kT}{m}\frac{\partial w}{\partial v}\right), \qquad \ldots\ldots(2.3)$$

where $V(x)$ is the potential energy of the particle, T the absolute temperature, and k Boltzmann's constant. The rate constant β, introduced in (2.2), determines the rate of approach to the equilibrium distribution in *velocity space*.

In most cases it is useful to carry out an exponential Fourier transformation of (2.3) with regard to the variable v. If we define

$$\Phi(x,u,t)=\int_{-\infty}^{\infty} w(x,v,t)\,e^{-iuv}\,dv, \qquad \ldots\ldots(2.4)$$

the derivative of w with regard to v and the product vw transform as

$$dw/dv \to iu\Phi; \quad vw \to i\partial\Phi/\partial u \qquad \ldots\ldots(2.5)$$

(cf. for example, Erdélyi 1954, chap. III). Equation (2.3) transforms into

$$\frac{\partial\Phi}{\partial t}+i\frac{\partial^2\Phi}{\partial x\partial u}-\frac{i}{m}\frac{\partial V}{\partial x}u\Phi=-\beta u\left(\frac{kT}{m}u\Phi+\frac{\partial\Phi}{\partial u}\right) \qquad \ldots\ldots(2.6)$$

and its equilibrium (Boltzmann) solution

$$w_0=\text{const.}\ \exp\left[-\{V(x)+\tfrac{1}{2}mv^2\}/kT\right] \qquad \ldots\ldots(2.7)$$

into

$$\Phi_0=\text{const.}\ \exp\left[-V/kT-kTu^2/2m\right]. \qquad \ldots\ldots(2.8)$$

Putting

$$\Phi=\Psi\exp(-kTu^2/2m) \qquad \ldots\ldots(2.9)$$

we obtain from (2.6)

$$\frac{D'\Psi}{D't}\equiv\frac{\partial\Psi}{\partial t}+i\frac{\partial^2\Psi}{\partial x\partial u}-\frac{iu}{m}\left[kT\frac{\partial\Psi}{\partial x}+\frac{\partial V}{\partial x}\Psi\right]=-\beta u\frac{\partial\Psi}{\partial u}. \qquad \ldots(2.10)$$

The use of the stochastic equations (2.6) or (2.10) rather than of (2.3) has two advantages. They contain only first order derivatives with regard to u, whereas (2.3) is of second order in v. Secondly, in the calculation of macroscopic averages any integration of w over the velocities v is accomplished in view of (2.4), if only Φ or Ψ are given the values corresponding to $u=0$. An expansion of Ψ as a power series in u is entirely equivalent to the more usual development of $w\exp(\frac{1}{2}mv/kT)$ into a series of Hermite polynomials; but in many cases (2.10) can be solved by other methods, which are not straightforward for the solution of equation (2.3).

For the large collisions considered in case (B) the right-hand side of the generalized Liouville equation contains an integral expression (Gross 1955 b), instead of the differential terms of (2.3), namely

$$\frac{Dw(x,v,t)}{Dt}=\beta_1\left(\frac{m}{2\pi kT}\right)^{1/2}\exp\left(\frac{-mv^2}{2kT}\right)\int w(x,v,t)\,dv-\beta_1 w, \qquad \ldots\ldots(2.11)$$

for which results for the transformed function

$$D'\Psi(x,u,t)/D't=\beta_1\Psi(x,0,t)-\beta_1\Psi(x,u,t) \qquad \ldots\ldots(2.12)$$

instead of (2.10). For the very large collisions of case (C) the equation becomes correspondingly

$$\frac{D'\Psi(x,u,t)}{D't}=-\beta_1\Psi(x,u,t)+\beta_1\Phi_0(x,0)\frac{\int\Psi(x,0,t)\,dx}{\int\Phi_0(x,0)\,dx}, \qquad \ldots\ldots(2.13)$$

Φ_0 being defined in (2.8). Gross (1955 b) has considered two additional cases; his equation (4*a*) corresponding to case (B) with an additional specular reflection, transforms into

$$D'\Psi(x,u,t)/D't=-\beta_1\Psi(x,u,t)+\beta_1(1-A)\,\Psi(x,-u,t)+\beta_1 A\Psi(x,0,t) \qquad \ldots\ldots(2.14)$$

and (4*d*), corresponding to partial coherence between the velocities before and after impact (cf. Keilson and Storer 1952), into

$$D'\Psi(x,u,t)/D't = -\beta_1\Psi(x,u,t) + \beta_1\Psi(x,\Gamma u,t). \qquad \ldots\ldots(2.15)$$

If the potential $V(x)$ is essentially harmonic, the equations can be further simplified if the Fourier transformation of (2.3) is performed over the variable x as well as over v.

All the above formulae are easily generalized for motion in several dimensions; thus for case (A), assuming β and m to be scalar constants, we obtain instead of (2.10)

$$\frac{D'\Psi}{D't} \equiv \frac{\partial\Psi}{\partial t} + i\sum\left[\frac{\partial^2\Psi}{\partial x_k \partial u_k} - \frac{kTu_k}{m}\frac{\partial\Psi}{\partial x_k} - \frac{u_k}{m}\frac{\partial V}{\partial x_k}\Psi\right] = -\beta\sum_k u_k\frac{\partial\Psi}{\partial u_k} \qquad \ldots\ldots(2.16)$$

with obvious analogies in the other cases.

It should be noted that whereas the rate constant β_1 in (2.11)–(2.15) represents the average collision rate, the damping parameter β in (2.2), (2.3), (2.6), (2.10) and (2.16) is equal to the product of this collision rate and the mean efficiency of each individual impact in destroying the coherence between the initial and final velocities; this second factor is equal to $1-\Gamma$ in the notation of (2.15) and becomes infinitesimally small in the limit of Brownian motion (cf. Keilson and Storer 1952).

§ 3. MODEL AND CALCULATIONS

In his first molecular theory of dielectric relaxation Debye (1913) considered a model in which polar molecules can rotate about fixed axes under the action of a local field and the randomizing influence of a non-polar substrate. He deduced a macroscopic behaviour in agreement with (1.1)–(1.6); the relaxation time τ in the case of rigid dipoles of moment of inertia I with negligible mutual interaction was found to be

$$\tau = \beta I/kT. \qquad \ldots\ldots(3.1)$$

Gross (1955 b) has treated the same model on the basis of the generalized Liouville equation for various types of collisions and demonstrated the deviation from pure Debye behaviour by expressing the solution in terms of continued fractions. The transformation (2.4), however, enables the same solutions to be expressed as analytical functions of the various parameters and the frequency ω or time t.

The position of the rotating dipole is completely specified by an angular variable ϕ which takes the place of x in (2.3)–(2.15); the velocity v now means $d\phi/dt$, and I replaces m. If an electric field F acts in the direction specified by $\phi=\pi/2$ the potential energy V becomes

$$V = -\mu F\sin\phi, \qquad \ldots\ldots(3.2)$$

μ being the component of the dipole moment normal to the axis of rotation.

The equilibrium solution Ψ_0 for the function Ψ (cf. (2.8)–(2.10)) is, apart from an irrelevant constant factor,

$$\Psi_0(\phi,u) = \exp(\mu F\sin\phi/kT) = 1 + \mu F\sin\phi/kT + \ldots, \qquad \ldots\ldots(3.3)$$

where the neglected terms in the expansion correspond to saturation effects with which we are not concerned here. If the function Ψ is expressed as a Fourier series in ϕ

$$\Psi(\phi,u,t) = \sum e^{ik\phi}\psi_k(u,t), \qquad \ldots\ldots(3.4)$$

the linear approximation to the stationary solution Ψ_0 involves only ψ_1 and ψ_{-1}, apart from the constant term $\psi_0 \equiv 1$; the same is found for the time dependent solution. In view of (2.10) and (3.2)

$$\psi_{-1}(u,t) = -\psi_1(-u,t). \qquad \ldots\ldots(3.5)$$

The total polarization of the system in the direction of the field is simply expressed as

$$P(t) = iN\mu[\psi_1(0,t) - \psi_{-1}(0,t)]/2\psi_0(0,t) \qquad \ldots\ldots(3.6)$$

where N is the number of dipoles per unit volume. It is sufficient to solve the stochastic equations either for the steady state response $P^*(\omega)\,e^{i\omega t}$ to an alternating field (1.3) or for the dielectric after-effect $P_a(t)$ in which a constant field F_0 is suddenly removed at $t=0$ after the stationary conditions (3.3) have been established, as the two solutions are related by the imaginary Laplace transformation

$$P^*(\omega) = P_0 - i\omega \int_0^\infty P_a(t)\,e^{-i\omega t}\,dt\,; \qquad \ldots\ldots(3.7)$$

in two of the three principal cases considered in § 2 it is easiest to derive analytical solutions for the after-effect.

For the Brownian motion type of fluctuation (case (A)) after the field has been removed, equations (2.10) and (3.4) lead to

$$\frac{\partial\psi_1}{\partial t} - \frac{\partial\psi_1}{\partial u} + \frac{ukT\psi_1}{I} = -\beta u \frac{\partial\psi_1}{\partial u} \qquad \ldots\ldots(3.8)$$

with the initial condition

$$\psi_1(u,0) = -i\mu F_0/2kT \qquad \ldots\ldots(3.9)$$

in view of (3.3). Substituting

$$\psi_1(u,t) = -(i\mu F_0/2kT)\exp[Z_1(u,t)], \qquad \ldots\ldots(3.10)$$

we obtain from (3.8) and (3.9)

$$\frac{\partial Z_1}{\partial t} + \frac{\partial Z_1}{du}(\beta u - 1) = -\frac{ukT}{I}, \qquad \ldots\ldots(3.11)$$

$$Z_1(u,0) = 0, \qquad \ldots\ldots(3.12)$$

of which the general and particular solutions are

$$Z_1 = -(kT/I\beta)(u+t) + f[(u-\beta^{-1})\,e^{-\beta t}] \qquad \ldots\ldots(3.13)$$

(f being any arbitrary function) and

$$Z_1 = -(kT/I\beta)[t + (u-\beta^{-1})(1-e^{-\beta t})] \qquad \ldots\ldots(3.14)$$

respectively. The decay of polarization P_a is thus obtained from (3.3), (3.5), (3.6), (3.10) and (3.14) as

$$P_a(t) = \tfrac{1}{2}(N\mu^2 F_0/kT)\exp[\gamma(1-\beta t - e^{-\beta t})] \qquad \ldots\ldots(3.15)$$

where the parameter γ (not to be confused with the quantity so denoted by Gross (1955 b), cf. (2.15)) is defined as

$$\gamma = kT/I\beta^2. \qquad \ldots\ldots(3.16)$$

The factor $\exp(-\gamma\,e^{-\beta t})$ can be expanded in powers of $e^{-\beta t}$ with the result

$$P_a(t)/P_0 = e^\gamma\left[\exp\left(\frac{-t}{\tau_0}\right) - \frac{\gamma}{1!}\exp\left(\frac{-t}{\tau_1}\right) + \frac{\gamma^2}{2!}\exp\left(\frac{-t}{\tau_2}\right) - \ldots\right], \qquad \ldots\ldots(3.17)$$

where the relaxation times τ_l are given by

$$1/\tau_l = (\gamma + l)\beta = kT/I\beta + l\beta. \qquad \ldots\ldots(3.18)$$

Each term in (3.17) corresponds to a purely exponential decay and for an alternating field transforms according to (1.6); hence

$$\frac{P^*(\omega)}{P} = e^{\gamma}\left[\frac{1}{1+i\omega\tau_0} - \frac{\gamma}{1\,!(1+i\omega\tau_1)} + \frac{\gamma^2}{2\,!(1+i\omega\tau_2)} - + \dots\right]. \qquad \dots\dots(3.19a)$$

In view of the definitions (3.18), the series (3.19*a*) can be expressed in terms of the confluent hypergeometric function ${}_1F_1\,(a\,;c\,;x)$, to which, in turn, Kummer's transformation can be applied (cf. for example Magnus and Oberhettinger 1949, p. 87), so that

$$\frac{P^*(\omega)}{P_0} = 1 - \frac{i\omega\tau_0}{1+i\omega\tau_0} e^{\gamma}\, {}_1F_1\left(\gamma + \frac{i\omega}{\beta}\,;\ \gamma+1+\frac{i\omega}{\beta}\,;\ -\gamma\right) \qquad \dots\dots(3.19b)$$

$$= 1 - \frac{i\omega\tau_0}{1+i\omega\tau_0}\, {}_1F_1\left(1;\ \ \gamma+1+\frac{i\omega}{\beta}\,;\ \gamma\right) \qquad \dots\dots(3.19c)$$

$$= 1 - \frac{i\omega\tau_0}{1+i\omega\tau_0}\left[1 + \frac{\gamma}{1+\gamma+i\omega/\beta} + \frac{\gamma^2}{(1+\gamma+i\omega/\beta)(2+\gamma+i\omega/\beta)} + \dots\right]. \qquad \dots\dots(3.19d)$$

A method of calculating $P^*(\omega)$ by solving, instead of (3.8), the equation

$$i\omega\psi_1 - \frac{d\psi_1}{du} + \frac{ukT\psi_1}{I} - \frac{iu\mu F}{2I} = -\beta u\frac{d\psi_1}{du} \qquad \dots\dots(3.20)$$

by means of an expansion of ψ_1 in powers of $(\beta u - 1)$ and application of (3.5) and (3.6) leads directly to (3.19*d*).

For the very large collisions of type (C) substitution of (3.4) in (2.13) leads to

$$\frac{\partial\psi_1}{\partial t} - \frac{\partial\psi_1}{\partial u} + \frac{ukT\psi_1}{I} = -\beta_1\psi_1 \qquad \dots\dots(3.21)$$

instead of (3.8). If the substitution (3.10) is carried out the particular solution of Z_1 is found to be

$$Z_1(u,t) = -\beta_1 t - \tfrac{1}{2}\gamma_1\beta_1^2 t^2 - \gamma_1\beta_1^2 tu\,;\quad \gamma_1 = kT/I\beta_1^2, \qquad \dots\dots(3.22)$$

and hence with the use of (3.6)

$$P_a(t)/P_0 = \exp\left(-\beta_1 t - \tfrac{1}{2}\gamma_1\beta_1^2 t^2\right), \qquad \dots\dots(3.23)$$

$$P^*(\omega)/P_0 = 1 - i\omega\int_0^\infty \exp\left[-(\beta_1 + i\omega)t - \tfrac{1}{2}\gamma_1\beta_1^2 t^2\right]dt \qquad \dots\dots(3.24a)$$

$$= 1 - (i\omega/\beta_1)(2/\gamma_1)^{1/2}\exp\zeta^2\,\operatorname{erfc}\zeta, \qquad \dots\dots(3.24b)$$

where erfc ζ denotes the complementary error function

$$\operatorname{erfc}\zeta = \int_\zeta^\infty \exp(-s^2)\,ds\,;\quad \zeta^2 = (1 + i\omega/\beta_1)^2/2\gamma_1. \qquad \dots\dots(3.25)$$

For large collisions of the type (B) an additional term $+\beta_1\psi_1(0,t)$ appears on the right-hand side of (3.21); the resulting equation can be integrated once to give

$$\psi_1(u,t) = -(iuF_0/2kT)\,\exp[Z_1(u,t)] + \beta_1\int_0^t \exp Z_1(u,\tau)\,\psi_1(0,t-\tau)\,d\tau \qquad \dots\dots(3.26)$$

where $Z_1(u,t)$ is the solution (3.22); equation (3.26) can be solved by Picard's method of successive approximations. For the dielectric after-effect $P_a(t)/P_0$ this yields an infinite series, the first term of which is the function (3.23) and each subsequent term is the convolution of the preceding term with this function, multiplied by a factor β_1. For the Laplace transformation (3.7), these convolutive powers are transformed into algebraic powers of the integral in (3.24*a*) (cf. Erdélyi 1954, chap. IV); the resulting geometric series is easily summed with the result

$$\frac{P^*(\omega)}{P_0}=1-\frac{i\omega}{\beta_1}\frac{\exp\zeta^2\operatorname{erfc}\zeta}{(\frac{1}{2}\gamma_1)^{1/2}-\exp\zeta^2\operatorname{erfc}\zeta} \qquad \ldots\ldots(3.27)$$

with the definitions of (3.25). The solutions (3.24)–(3.27) have already been obtained by Gross (1955 b), but the present derivation appears simpler.

§ 4. RESULTS AND DISCUSSION

Generalized Brownian Motion

The dielectric properties of a system of dipoles rotating freely about fixed axes under the influence of the Brownian motion due to a viscous substrate are described by equations (3.15)–(3.19); (3.15) and (3.19*a*) have previously been published by the writer without derivation (Sack 1953 b). The factor ½ occurring in (3.15) applies to a model in which all the axes of rotation are parallel; if they are randomly oriented in space, the corresponding factor is ⅓. Formula (3.15) shows that the decay of polarization in the dielectric after-effect, for periods $t \gg \beta^{-1}$, follows an exponential law with the Debye relaxation time τ_0 given in (3.1), but that the initial rate of decay immediately after the switching-off of the field is zero since

$$\exp\left[\gamma(1-\beta t-e^{-\beta t})\right]=1-\frac{1}{2}\gamma(\beta t)^2+\frac{1}{6}\gamma(\beta t)^3-\left(\frac{1}{24}\gamma-\frac{1}{8}\gamma^2\right)(\beta t)^4-+\ldots; \qquad \ldots\ldots(4.1)$$

this is in agreement with the general remarks of § 1. It is interesting to note that the exponent in (3.15) shows the same time dependence, except for sign, as the mean square displacement of a particle subject to translational Brownian motion from a given initial position according to the Fokker–Planck equation (Ornstein 1919, Fürth 1920)

$$\langle\delta x^2\rangle=(2kT/m\beta^2)(\beta t-1+e^{-\beta t}), \qquad \ldots\ldots(4.2)$$

whereas the older Einstein expression

$$\langle\delta x^2\rangle=2kTt/m\beta \qquad \ldots\ldots(4.3)$$

bears the same relation to the Debye relaxation formula (1.1) and (3.1). This analogy, however, seems to apply only to the specific model considered; no corresponding expression could be found for free rotation in space.

Equations (3.17)–(3.19*a*) show that although a single microscopic fluctuation mechanism is operative, a macroscopic description in terms of the usual relaxation concepts would involve an infinity of discrete relaxation times, whose reciprocals form an arithmetic progression; the corresponding amplitudes have alternate signs and satisfy the condition (1.10). The leading longest relaxation time τ_0 has the same value as the single relaxation time arising out of the Debye approximation (3.1), but the corresponding amplitude is greater by a factor e^{γ}.

An alternative expression to (3.19) for the frequency response $P^*(\omega)$ of the polarization has been given by Gross (1955 b) in terms of a continued fraction. In the notation of the present paper this can be written as

$$\frac{P^*(\omega)}{P_0}=1-\frac{i\omega}{\beta}\left[\frac{1\,|}{|\,i\omega/\beta}+\frac{\gamma\,|}{|\,1+i\omega/\beta}+\frac{2\gamma\,|}{|\,2+i\omega/\beta}+\frac{3\gamma\,|}{|\,3+i\omega/\beta}+\ldots\right] \qquad \ldots\ldots(4.4)$$

and could have been derived from a power series development of the function ψ_1 in (3.20). The conclusions to be drawn from the solutions (3.19) and (4.4) are essentially the same. The second approximation of (4.4) simplifies to

$$\frac{P^*(\omega)}{P_0}=\frac{1}{1+i\omega\tau_0-\omega^2\tau_0/\beta}, \qquad \ldots\ldots(4.5)$$

and the third approximation of (3.19*d*) to

$$\frac{P^*(\omega)}{P_0}=\frac{1}{1+i\omega\tau_0}\cdot\frac{1+i\omega\beta\tau_1\tau_2}{(1+i\omega\tau_1)(1+i\omega\tau_2)}. \qquad \ldots\ldots(4.6)$$

Both approximations are valid whenever $\gamma \ll 1$; this condition implies, in view of the definition (3.16), that β is much larger than $(kT/I)^{1/2}$, which is the mean thermal angular velocity, in other words that the damping is so strong that the equilibrium distribution in velocity is reached before the dipole has had time to change its direction appreciably. In this case (4.5) and (4.6) reduce to the Debye formula (1.6) provided $\omega\tau_1 \sim \omega/\beta \ll 1$, but show clearly the more rapid decrease of absorption at very high frequencies. Debye himself (1934) has indicated that his equations (1.6) are no longer applicable at frequencies approaching the mean collision rate, but in view of the comments at the end of § 2, for small (Brownian) collisions, the rate constant β is only a small fraction of the collision rate, and hence Debye's criterion is not stringent enough. The breakdown of the Debye equations at frequencies at which ω is of the same order as or larger than β is in agreement with the general theory of Brownian motion, according to which the Smoluchowski equation (on which Debye's treatment is essentially based) gives results consistent with the more accurate Kramers equation only for periods t large compared with β^{-1} (Chandrasekhar 1943).

For larger values of γ, the higher terms in the expansion (3.19*a*) become progressively more important, and the deviation from Debye's results extends over the whole frequency range, in the sense that the absorption–frequency curve becomes narrower. But, as pointed out by Gross (1955 b), in this case the classical model is no longer a good approximation in view of the discrete spacing of rotational levels, and a quantum-mechanical treatment must be used.

Abrupt Collisions

The behaviour of the dielectric models with large collisions (types (B) and (C)) has already been discussed by Gross (1955 b) on the basis of the solutions (3.24*b*), (3.25) and (3.27) as well as of several expansions in terms of continued fractions. For the very large collisions of type (C) equation (3.23) shows that the dielectric after-effect in its initial stages follows an exponential law with a relaxation time $\tau=\beta_1^{-1}$, but that eventually the decay will be considerably more rapid. If, however, γ_1 is small, most of the polarization will have decayed before the deviation from a pure exponential law becomes appreciable. From the fact that for any positive value of γ_1 the expression (3.24*a*) taken as a function of ω has no singularities, except at infinity, one can deduce that the behaviour of the material cannot be described by any distribution of relaxation times. The use of the asymptotic expansion (Magnus and Oberhettinger 1949, p. 96)

$$2\zeta \exp \zeta^2 \operatorname{erfc} \zeta=1-\frac{1}{2}\zeta^2+\frac{1.3}{4}\zeta^4-\frac{1.3.5}{8}\zeta^6+-\ldots, \qquad \ldots(4.7)$$

valid for large ζ, i.e. small γ_1, enables us to re-write (3.24*b*) and (3.25) in the form

$$\frac{P^*(\omega)}{P_0}=\frac{1}{1+i\omega/\beta_1}-\frac{i\omega\gamma_1}{(1+i\omega/\beta_1)^3}+\frac{1.3i\omega\gamma_1^2}{(1+i\omega/\beta_1)^5}-+\ldots \qquad \ldots\ldots(4.8)$$

which again shows that the material behaves approximately as a Debye dielectric if $\gamma_1 \ll 1$. The arguments of § 1 do not apply to this model and the approximate Debye response persists up to the highest frequencies.

For large collisions of the type (B) which leave the direction of the dipole unchanged, any description in terms of a distribution of relaxation times analogous to (3.19*a*) would involve values of $\tau = \lambda^{-1} = i/\omega$ for which the denominator in (3.27) vanishes. There is only one such value of λ on the real axis; for small values of γ_1 the relevant equation can be obtained from the expansion (4.7) as

$$\frac{\lambda}{\beta_1} = \frac{\gamma_1}{(1-\lambda/\beta_1)^2} - \frac{3\gamma_1^2}{(1-\lambda/\beta_1)^4} + \frac{1.3.5\gamma_1^3}{(1-\lambda/\beta_1)^6} - + \dots \qquad \dots\dots(4.9)$$

with the solution

$$\lambda = \tau^{-1} = \beta_1(\gamma_1 - \gamma_1^2 + 4\gamma_1^3 - + \dots). \qquad \dots\dots(4.10)$$

If β and γ are substituted for β_1 and γ_1, this relaxation time agrees with the Debye value (3.1) for the Brownian collision mechanism only to the first power in γ. The two models agree well at very high frequencies or very short periods, for a power series development of the decay function $P_a(t)/P_0$ as outlined after (3.26) agrees with (4.1) up to the term in t^4. The same agreement is also shown up by Gross's solution for $P^*(\omega)$ as a continued fraction which is

$$\frac{P^*(\omega)}{P_0} = 1 - \frac{i\omega}{\beta}\left[\frac{1\,|}{|\,i\omega/\beta} + \frac{\gamma\,|}{|\,1+i\omega/\beta} + \frac{2\gamma\,|}{|\,1+i\omega/\beta} + \frac{3\gamma\,|}{|\,1+i\omega/\beta} + \dots\right]; \qquad \dots\dots(4.11)$$

like (4.4), this simplifies to (4.5) in second approximation and approximates (4.4) more and more closely as ω increases. As (4.7) is not a convergent but only an asymptotic expansion, the same is true of (4.8) and (4.9), and presumably also of (4.10) and (4.11). These formulae cannot, therefore, be used to determine the behaviour of the model for large values of γ (or γ_1); but for such systems, as pointed out above, the classical approximation is not adequate in any case.

For a discussion of the applicability of the various models to the dielectric absorption of compressed gases the reader is referred to the paper by Gross (1955 b).

ACKNOWLEDGMENTS

This work formed part of the fundamental research programme of the British Rayon Research Association. The writer wishes to thank Professor H. Fröhlich, Professor J. Meixner, Dr. B. Gross, Dr. L. R. G. Treloar, Dr. H. Pelzer and Dr. A. S. Lodge for stimulating discussions and criticism.

REFERENCES

CHANDRASEKHAR, S., 1943, *Rev. Mod. Phys.*, **15**, 1.
DEBYE, P., 1913, *Ber. Dtsch. Phys. Ges.*, **15**, 777; 1934, *Phys. Z.*, **35**, 101; 1945, *Polar Molecules* (New York: Dover Publications).
DMITRIEV, V. A., and GUREVICH, C. B., 1946, *Zh. Eksp. Teor. Fiz.*, **16**, 937.
ERDÉLYI, A., (Ed.), 1954, *Tables of Integral Transforms*, Vol. I. (Bateman Manuscript Project; New York: McGraw-Hill).
FRÖHLICH, H., 1946, *Nature, Lond.*, **157**, 478.
FRÖHLICH, H., and PLATZMAN, R. L., 1953, *Phys. Rev.*, **92**, 1152.
FÜRTH, R., 1920, *Z. Phys.*, **2**, 244.
GROSS, B., 1953, *Kolloid Z.*, **134**, 73, 83.
GROSS, E. P., 1955 a, *Phys. Rev.*, **97**, 395; 1955 b, *J. Chem. Phys.*, **23**, 1415.
HARPER, P. G., and O'DWYER, J. J., 1955, *Proc. Phys. Soc.* A, **68**, 1184.

Keilson, J., and Storer, J. E., 1952, *Quart. Appl. Maths.*, **10**, 243.

Kramers, H. A., 1940, *Physica*, **7**, 284.

Magnus, W., and Oberhettinger, F., 1949, *Formulas and Theorems for the Special Functions of Mathematical Physics* (New York : Chelsea Publishing Co.).

Ornstein, L. S., 1919, *Proc. Acad. Amsterdam*, **21**, 96.

Powles, J. G., 1948, *Trans. Faraday Soc.*, **44**, 802.

Rocard, M. Y., 1933, *J. Phys. Radium*, **4**, 247.

Sack, R. A., 1953 a, *Kolloid Z.*, **134**, 16; 1953 b, *Ibid.*, **134**, 83; 1957, *Proc. Phys. Soc.* B, **70**.

Saito, N., and Kato, T., 1956, *Busseiron Kenkyu*, **92**, 103.

Relaxation Processes and Inertial Effects
II: Free Rotation in Space

By R. A. SACK†

Department of Theoretical Physics, University of Liverpool
and
British Rayon Research Association, Wythenshawe, Manchester

MS. received 20th November 1956

Abstract. The calculations of the effect of inertia on the dielectric properties of systems of identical non-interacting rigid dipoles with various types of fluctuation are extended to the case of free rotation in space. The equations of motion, and hence the stochastic equations, depend on all three moments of inertia; the dumb-bell and the spherical top models are treated in detail. The probability distribution in phase space is conveniently replaced by a distribution in configuration–velocity space which simplifies the fluctuation terms, but necessitates a modification of the Liouville equation. The response of the polarization to an alternating electric field can be expressed in terms of continued fractions similar to the solution obtained by E. P. Gross for the plane rotation model. For all collision mechanisms which leave the orientation of the dipoles unaltered the second approximation yields a modification of Debye's relaxation formula which has previously been derived by Powles and by Gross; the precise nature of the model affects the result only in higher approximation. The relation of the exact solutions and several approximate formulae is discussed and a misprint in a formula by Rocard is pointed out. For fluctuations of the Brownian movement type the frequency response of the dumb-bell model tends towards that of a damped harmonic oscillator as the viscosity decreases. For abrupt collisions the results can be expressed in an analytic form.

§ 1. INTRODUCTION

In the first paper of this series (Sack 1957, to be referred to as I) the dielectric behaviour at high frequencies was calculated for a system of rigid dipoles rotating about fixed axes under the influence of an orienting electric field and several types of frictional interaction with a non-polar background. As shown by Gross (1955) such a model is particularly relevant for the description of the properties of compressed polar gases at microwave frequencies. For strong damping the results agree with the well-known Debye formula at not too high frequencies; in the limit of very high frequencies, however, the product of loss factor and frequency tends to zero, whereas Debye's theory predicts a monotonic increase of this product towards a finite limit. The reason for the failure of the Debye theory at very high frequencies is that it is based on Smoluchowski's diffusion equation which assumes dynamic equilibrium between the orienting and viscous forces to be established at all times; for very short periods this assumption is not justified. A more consistent treatment should be based on the generalized Liouville equation in which both positions and momenta (or velocities) appear as independent variables; for the model with plane rotation exact formulae have been derived on this basis independently by Gross (1955) and by the writer (Sack 1953, 1957).

---

† Now at Department of Theoretical Chemistry, University of Cambridge.

A model which is more satisfactory for the description of actual gases or fluids is a system of dipoles which can rotate freely in space; in the Debye approximation the two systems behave alike, except for the magnitude of the relaxation times which for identical moments of inertia and frictional constants differ by a factor 2. In the present paper the time dependence of the macroscopic polarization of a system of non-interacting rigid dipoles rotating freely in space in a non-polar heat bath will be derived on the basis of the classical Liouville equation.

A number of possible mechanisms of collisions with the background have been discussed by Gross (1955) and in I; they are

(A) Brownian motion interaction where each collision leaves the orientations of the molecules unaltered and changes their (angular) velocities by infinitesimal amounts only.

(B) Large collisions by which the orientations remain unaffected, but the velocity distribution after impact is Maxwellian regardless of the initial velocity.

(C) Very large collisions after which the dipoles can be found anywhere in phase space with a probability proportional to the Boltzmann distribution corresponding to the instantaneous Hamiltonian of the system.

(D) Intermediate collisions which leave the orientation of the molecules unaltered, but preserve some correlation between the initial and final velocities; this mechanism has been suggested by Kellson and Storer (1952); it includes both (A) and (B) as limiting cases.

In all cases it is assumed that the collisions are adiabatic, i.e. can be treated as instantaneous, and that no correlation exists between successive collisions.

In the Debye approximation the relaxation time calculated for molecules whose dipole directions coincide with a principal axis of inertia is independent of the moment of inertia I_z about this axis. The equations of motion, however, and hence the Liouville equation contain I_z explicitly, and different results are to be expected as I_z is varied. The present investigation is confined to the two most important limiting cases: the dumb-bell or needle model, for which $I_z=0$, and the rotating sphere. In all cases the appropriate coordinate systems are curvilinear, and it has been found that in the presence of fluctuation terms the probability distribution is more conveniently referred to the configuration–velocity space than to the phase space. The stochastic equations for the various collision mechanisms are set up and solved for the dumb-bell model in § 2; the corresponding calculations for the spherical model are given in § 3. In § 4 the results are discussed and compared with approximate solutions derived by other workers in this field.

§ 2. The Rotating Needle Model

The Debye theory of dielectric relaxation for a system of identical rigid spherical dipoles with negligible mutual interaction and free rotation except for Brownian motion and viscous damping by a non-polar substrate predicts a relaxation time τ_8

$$\tau_8=\beta I/2kT \qquad \ldots\ldots(2.1)$$

where I is the moment of inertia of the sphere, T the absolute temperature, k Boltzmann's constant and β a viscous parameter relating the angular velocity $\mathbf{v}$ with the angular acceleration (in the absence of orienting forces)

$$d\mathbf{v}/dt = -\beta\mathbf{v} \qquad \ldots\ldots(2.2)$$

(Debye 1945). The result (2.1) applies to all molecules for which the dipole direction constitutes an axis of symmetry, provided I and β refer to rotations about an axis normal to the figure axis.

If the influence of inertia is to be taken into account, the simplest model to be considered is that of a system of freely rotating needles (dumb-bells) the orientation of which is completely specified by the spherical polar angles θ and ϕ referred to the fixed direction of an applied electric field as polar axis. The total energy (Hamiltonian) H of a rotating dipole of moment μ is

$$H = \tfrac{1}{2}I(v_\theta{}^2 + v_\phi{}^2) - \mu F\cos\theta = (p_\theta{}^2 + p_\phi{}^2/\sin^2\theta)/2I - \mu F\cos\theta \quad \ldots\ldots(2.3)$$

where the (angular) velocity components v_θ and v_ϕ and the momenta p_θ and p_ϕ are defined by

$$p_\theta = Iv_\theta = I\,d\theta/dt\,; \qquad p_\phi = I\sin\theta v_\phi = I\sin^2\theta\,d\phi/dt. \quad \ldots\ldots(2.4)$$

The distribution function $W(\theta, \phi, p_\theta, p_\phi, t)$ in phase space for an assembly of such rotators can, like H, be assumed independent of ϕ and thus the Liouville equation

$$\frac{DW}{Dt} \equiv \frac{\partial W}{\partial t} + \sum_i \left(\frac{\partial W}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial W}{\partial p_i}\frac{\partial H}{\partial q_i}\right) = 0 \qquad \ldots\ldots(2.5)$$

becomes, in view of (2.3),

$$\frac{DW}{Dt} \equiv \frac{\partial W}{\partial t} + \frac{\partial W}{\partial \theta}\frac{p_\theta}{I} + \frac{\partial W}{\partial p_\theta}\left(\frac{\cos\theta p_\phi{}^2}{\sin^3\theta I} - \mu F\sin\theta\right) = 0. \quad \ldots\ldots(2.6)$$

If Brownian motion comes into play, the fluctuation terms complementing the Liouville equation become very complicated for curvilinear canonical variables (Green 1952); it is, therefore, an advantage to use a distribution in configuration–velocity space for which the fluctuation terms are given by the Fokker–Planck equation (cf. Chandrasekhar 1943). If we define a density function $w(\theta, \phi, v_\theta, v_\phi, t)$ such that the probability dP of finding a dipole in the four-dimensional space element described by the infinitesimal increments $d\theta$, $d\phi$, dv_θ, dv_ϕ is given by

$$dP = w(\theta, \phi, v_\theta, v_\phi, t)\sin\theta\,d\theta\,d\phi\,dv_\theta\,dv_\phi, \qquad \ldots\ldots(2.7)$$

the relations between W and w and their derivatives at a given point in phase space become, because of (2.4),

$$W = \frac{w}{I^2},\ \frac{\partial W}{\partial \theta} = \frac{1}{I^2}\left(\frac{\partial w}{\partial \theta} - v_\phi\cot\theta\frac{\partial w}{\partial v_\phi}\right)\,; \quad \frac{\partial W}{\partial p_\theta} = \frac{1}{I^3}\frac{\partial w}{\partial v_\theta}\,; \quad \ldots\ldots(2.8)$$

the difference in the expression for the θ-derivatives arises from the fact that in one case p_ϕ, and in the other v_ϕ, is kept constant. The expression (2.6) transforms into

$$\frac{Dw}{Dt} \equiv \frac{\partial w}{\partial t} + v_\theta\frac{\partial w}{\partial \theta} + \cot\theta v_\phi\left(v_\phi\frac{\partial w}{\partial v_\theta} - v_\theta\frac{\partial w}{\partial v_\phi}\right) - \frac{\mu F}{I}\sin\theta\frac{\partial w}{\partial v^\theta}, \quad \ldots\ldots(2.9)$$

$$\partial w/\partial\phi = 0. \qquad \ldots\ldots(2.10)$$

Following I we carry out a Fourier transformation over the velocities; if we define

$$\Phi(\theta, u_\theta, u_\phi) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(\theta, v_\theta, v_\phi) \exp(-iu_\theta v_\theta - iu_\phi v_\phi)\, dv_\theta\, dv_\phi \quad \ldots\ldots(2.11)$$

and further substitute

$$\Phi(\theta, u_\theta, u_\phi) = \Psi(\theta, u_\theta, u_\phi) \exp[-(u_\theta^2 + u_\phi^2)kT/2I], \quad \ldots\ldots(2.12)$$

(2.9) transforms into

$$\begin{aligned}\frac{D'\Psi}{D't} \equiv \frac{\partial\Psi}{\partial t} + i\frac{\partial^2\Psi}{\partial\theta\partial u_\theta} - \frac{iu_\theta}{I}\left(kT\frac{\partial\Psi}{\partial\theta} + \mu F \sin\theta\Psi\right) \\ + i\cot\theta\left(\frac{\partial}{\partial u_\phi} - \frac{kT}{I}u_\phi\right)\left(u_\phi\frac{\partial\Psi}{\partial u_\theta} - u_\theta\frac{\partial\Psi}{\partial u_\phi}\right).\end{aligned} \quad \ldots\ldots(2.13)$$

For any field F the equilibrium solution Ψ_0 is given by

$$4\pi\Psi_0(\theta, u_\theta, u_\phi; F) = 1 + \mu F\cos\theta/kT + \ldots, \quad \ldots\ldots(2.14)$$

independently of u_θ and u_ϕ; terms proportional to higher powers of $\mu F/kT$ correspond to saturation effects and will be neglected in all cases.

The fluctuation terms in the equations for Ψ for the various collision mechanisms described in § 1 have been derived for the one-dimensional case in I and by Keilson and Storer (1952) for mechanism (D). As in the present case the derivations are entirely analogous, only the results need be quoted. One finds

$$\left.\begin{aligned} & = -\beta(u_\theta\partial\Psi/\partial u_\theta + u_\phi\partial\Psi/\partial u_\phi), && \text{(A)} \\ \frac{D'\Psi(\theta, u_\theta, u_\phi, t)}{D't} & = -\beta_1\Psi(\theta, u_\theta, u_\phi, t) + \beta_1\Psi(\theta, 0, 0, t), && \text{(B)} \\ & = -\beta_1\Psi(\theta, u_\theta, u_\phi, t) + \beta_1\Psi_0(\theta; F), && \text{(C)} \\ & = -\beta_1\Psi(\theta, u_\theta, u_\phi, t) + \beta_1\Psi(\theta, \Gamma u_\theta, \Gamma u_\phi, t). && \text{(D)} \end{aligned}\right\} \quad \ldots\ldots(2.15)$$

The rate constant β in (A) is the same as in (2.2), in (B), (C) and (D) β_1 denotes the mean collision rate; the function Ψ_0 in (C) has been defined in (2.14) and the parameter Γ in (D) is a measure of the correlation of the velocities before and after impact.

It was shown in I that a Fourier transformation of the one-dimensional Fokker–Planck–Kramers equation resulted in a differential equation which was only of first order in the variable u. In the present case this advantage appears to be lost as (2.13) contains second-order derivatives in u_θ and u_ϕ. However, by means of a transformation to polar coordinates in u-space and a separation of the angular variables we can simplify (2.13) to a form involving only first-order derivatives in u. For this purpose we put

$$u_\theta = u\cos\phi_u; \qquad u_\phi = u\sin\phi_u; \quad \ldots\ldots(2.16)$$

we find that Ψ can be expressed by

$$4\pi\Psi = 1 + \cos\theta\psi_1(u, t) + \sin\theta\cos\phi_u\psi_2(u, t) + \ldots, \quad \ldots\ldots(2.17)$$

where ψ_1 is an even and ψ_2 an odd function of u, both of the order $\mu F/kT$, and higher terms are again neglected. Substitution of (2.16) and (2.17) into (2.13) and (2.15) after some elementary but lengthy calculations leads to

$$\frac{D'\Psi}{D't} \equiv \cos\theta\left(\frac{\partial\psi_1}{\partial t}+i\frac{\partial\psi_2}{\partial u}+i\frac{\psi_2}{u}-i\frac{ukT\psi_2}{I}\right)$$
$$+\sin\theta\cos\phi_u\left(\frac{\partial\psi_2}{\partial t}-i\frac{\partial\psi_1}{\partial u}+i\frac{kTw\psi_1}{I}-i\frac{u\mu F}{I}\right) \quad \ldots\ldots(2.18)$$

and

$$\left.\begin{aligned} \frac{D'\Psi}{D't} &= -\beta u\left(\cos\theta\frac{\partial\psi_1}{\partial u}+\sin\theta\cos\phi_u\frac{\partial\psi_2}{\partial u}\right), && \text{(A)}\\ &= \beta_1\cos\theta[\psi_1(0)-\psi_1(u)]-\beta_1\sin\theta\cos\phi_u\psi_2(u), && \text{(B)}\\ &= \beta_1\cos\theta[\mu F/kT-\psi_1(u)]-\beta_1\sin\theta\cos\phi_u\psi_2(u), && \text{(C)}\\ &= \beta_1\cos\theta[\psi_1(\Gamma u)-\psi_1(u)]+\beta_1\sin\theta\cos\phi_u[\psi_2(\Gamma u)-\psi_2(u)]. && \text{(D)} \end{aligned}\right\} \quad \ldots(2.19)$$

The macroscopically important quantity is the polarization P of the system, which in view of (2.11), (2.12) and (2.17) is given by

$$P(t)=N\mu\iiiint w\cos\theta\sin\theta\,d\theta\,d\phi\,dv_\theta\,dv_\phi=\tfrac{1}{3}N\mu\psi_1(0,t) \quad \ldots\ldots(2.20)$$

where N is the number of dipoles per unit volume; for a static field F_0, (2.14) and (2.20) yield

$$P_0=N\mu^2F_0/3kT. \quad \ldots\ldots(2.21)$$

To calculate the steady state response $P(t)=P^*(\omega)e^{i\omega t}$ to an alternating field $F=F_0e^{i\omega t}$ for the Brownian type of collisions (case (A)) we expand

$$\left.\begin{aligned} \psi_1&=(\mu F_0/kT)e^{i\omega t}(1-a_0-a_2(\beta u)^2-a_4(\beta u)^4-\ldots),\\ \psi_2&=(i\mu F_0/kT)e^{i\omega t}(b_1\beta u+b_3(\beta u)^3+b_5(\beta u)^5+\ldots), \end{aligned}\right\} \quad \ldots\ldots(2.22)$$

where the coefficients a_n, b_n depend on frequency only. For $P^*(\omega)$ we obtain from (2.20) and (2.22)

$$P^*(\omega)=P_0(1-a_0). \quad \ldots\ldots(2.23)$$

Substitution of (2.22) into (2.18) and (2.19(A)) leads to the following recurrence relations

$$\left.\begin{aligned} &i\omega'a_0+2b_1=i\omega',\\ &(i\omega'+n)a_n+(n+2)b_{n+1}-\gamma b_{n-1}=0, \quad (n>0)\\ &(i\omega'+n)b_n+(n+1)a_{n+1}-\gamma a_{n-1}=0, \end{aligned}\right\} \quad \ldots\ldots(2.24)$$

where

$$\omega'=\omega/\beta, \qquad \gamma=kT/I\beta^2. \quad \ldots\ldots(2.25)$$

This set of three-term recurrence relations can be solved for a_0 in terms of a continued fraction (Perron 1950); the solution of (2.23) is found to be

$$\frac{P^*(\omega)}{P_0}=1-\frac{i\omega'\,|}{|\,i\omega'}+\frac{2\gamma\,|}{|\,1+i\omega'}+\frac{2\gamma\,|}{|\,2+i\omega'}+\frac{4\gamma\,|}{|\,3+i\omega'}+\frac{4\gamma\,|}{|\,4+i\omega'}+\frac{6\gamma\,|}{|\,5+i\omega'}+\ldots. \quad \ldots\ldots(2.26)$$

If the other collision mechanisms are operative, the expansion (2.22) should be made in powers of $(\beta_1 u)$; with the abbreviations

$$\omega''=\omega/\beta_1, \qquad \gamma_1=kT/I\beta_1^2, \quad \ldots\ldots(2.27)$$

the recurrence relations become for mechanism (D)

$$\left.\begin{aligned} &i\omega''a_0+2b_1=i\omega'',\\ &(i\omega''+1-\Gamma^n)a_n+(n+2)b_{n+1}-\gamma_1 b_{n-1}=0, \quad (n>0)\\ &(i\omega''+1-\Gamma^n)b_n+(n+1)a_{n+1}-\gamma_1 a_{n-1}=0, \end{aligned}\right\} \quad \ldots\ldots(2.28)$$

with the solution

$$\frac{P^*(\omega)}{P_0}=1-\frac{i\omega''|}{|i\omega''}+\frac{2\gamma_1\;|}{|1-\Gamma+i\omega''}+\frac{2\gamma_1\;|}{|1-\Gamma^2+i\omega''}$$
$$+\frac{4\gamma_1\;|}{|1-\Gamma^3+i\omega''}+\frac{4\gamma_1\;|}{|1-\Gamma^4+i\omega''}+\ldots. \qquad \ldots\ldots(2.29)$$

Case (B) results from (D) if $\Gamma=0$; the polarization $P^*(\omega)$ in case (B) is therefore

$$\frac{P^*(\omega)}{P_0}=1-\frac{i\omega''|}{|i\omega''}+\frac{2\gamma_1\;|}{|1+i\omega''}+\frac{2\gamma_1\;|}{|1+i\omega''}+\frac{4\gamma_1\;|}{|1+i\omega''}+\ldots. \qquad \ldots\ldots(2.30)$$

The recurrence relations for mechanism (C) are the same as for (B) with the exception of the first, which becomes

$$(1+i\omega'')a_0+2b_1=i\omega'', \qquad \ldots\ldots(2.31)$$

the solution being given by

$$\frac{P^*(\omega)}{P_0}=1-\frac{i\omega''\;|}{|1+i\omega''}+\frac{2\gamma_1\;|}{|1+i\omega''}+\frac{2\gamma_1\;|}{|1+i\omega''}+\frac{4\gamma_1\;|}{|1+i\omega''}+\ldots. \qquad \ldots\ldots(2.32)$$

The frequency response for this case (C) can also be expressed in an analytic form involving the exponential integral of complex argument; the result, derived in the Appendix, is

$$(1+i\omega'')P^*(\omega)/P_0=1+i\omega''X[1+\mathrm{e}^X\mathrm{Ei}(-X)] \qquad \ldots\ldots(2.33)$$

where

$$X=(1+i\omega'')^2/2\gamma_1; \quad -\mathrm{Ei}(-X)=\int_X^\infty \mathrm{e}^{-s}\,ds/s. \qquad \ldots\ldots(2.34)$$

The ratios b_1/a_0 for $\Gamma=0$ are identical in (2.28) and (2.31) since all the later recurrence relations are the same; if this ratio is calculated from (2.23), (2.31) and (2.33) and substituted in (2.28), the analytic solution for case (B) becomes, with the abbreviations (2.27) and (2.34),

$$\frac{P^*(\omega)}{P_0}=\frac{1+i\omega'')X[1+\mathrm{e}^X\mathrm{Ei}(-X)]}{X[1+\mathrm{e}^X\mathrm{Ei}(-X)]+i\omega''}. \qquad \ldots\ldots(2.35)$$

§ 3. THE ROTATING SPHERE MODEL

The orientation of a rigid polar molecule of spherical shape is best specified by the Euler angles θ, ϕ, χ where θ and ϕ indicate the direction of the dipole with respect to the fixed direction of the field, as in § 2, and χ is the angle between a given equatorial line (i.e. a line normal to the dipole) which rotates with the sphere and the nodal line; this latter is defined as the momentary intersection of the equatorial plane and the plane containing the field and dipole directions (cf. Slater and Frank 1947). The angular velocity $\mathbf{v}$ is usually resolved into three components v_x, v_y, v_z about orthogonal axes fixed within the molecule. In view of the axial symmetry of the problem it is simpler to use components v_s about the dipole direction, v_θ about the nodal line, and v_ϕ about the equatorial line perpendicular to the nodal line. These angular velocity components are related to the time derivatives of the Euler angles by

$$v_\theta=\dot\theta, \quad v_\phi=\sin\theta\dot\phi, \quad v_s=\dot\chi+\cos\theta\dot\phi. \qquad \ldots\ldots(3.1)$$

From the expression for the kinetic energy K for a rotating sphere

$$K=\tfrac{1}{2}I(v_\theta^2+v_\phi^2+v_z^2)=\tfrac{1}{2}I(\dot\theta^2+\dot\phi^2+\dot\chi^2+2\cos\theta\dot\phi\dot\chi) \qquad \ldots\ldots(3.2)$$

the momenta are defined as

$$\left.\begin{aligned} p_\theta&=\partial K/\partial\dot\theta=I\dot\theta=Iv_\theta,\\ p_\phi&=\partial K/\partial\dot\phi=I(\dot\phi+\dot\chi\cos\theta)=I(v_\phi\sin\theta+v_z\cos\theta),\\ p_\chi&=\partial K/\partial\dot\chi=I(\dot\chi+\dot\phi\cos\theta)=Iv_z;\end{aligned}\right\} \qquad \ldots\ldots(3.3)$$

if K is expressed in terms of these momenta, the Hamiltonian H becomes

$$H=(2I)^{-1}[p_\theta^2+(p_\phi^2+p_\chi^2-2p_\phi p_\chi\cos\theta)/\sin^2\theta]-\mu F\cos\theta. \qquad \ldots\ldots(3.4)$$

The probability distribution W in the six-dimensional phase space will, like H, be independent of both ϕ and χ. Liouville's equation (2.5) becomes accordingly

$$\frac{DW}{Dt}\equiv\frac{\partial W}{\partial t}+\frac{p_\theta}{I}\frac{\partial W}{\partial\theta}+\frac{\partial W}{\partial p_\theta}\left[\frac{\cot\theta}{I\sin^2\theta}(p_\phi^2+p_\chi^2-2p_\phi p_\chi\cos\theta)\right.$$
$$\left.-\frac{p_\phi p_\chi}{I\sin\theta}-\mu F\sin\theta\right]=0. \qquad \ldots\ldots(3.5)$$

A new probability density w in configuration–velocity space is introduced with the probability dP defined in analogy to (2.7) as

$$dP=w\sin\theta\,d\theta\,d\phi\,d\chi\,dv_\theta\,dv_\phi\,dv_z; \qquad \ldots\ldots(3.6)$$

the formulae (3.3) and (3.6) lead to the relations between W and w and their derivatives

$$W=\frac{w}{I^3};\quad \frac{\partial W}{\partial\theta}=\frac{1}{I^3}\left(\frac{\partial w}{\partial\theta}-v_\phi\cot\theta\frac{\partial w}{\partial v_\phi}+v_z\frac{\partial w}{\partial v_\phi}\right);\quad \frac{\partial W}{\partial p_\theta}=\frac{1}{I^4}\frac{\partial w}{\partial v_\theta}. \qquad \ldots\ldots(3.7)$$

The Liouville expression (3.5) becomes with (3.3) and (3.7)

$$\frac{Dw}{Dt}\equiv\frac{\partial w}{\partial t}+v_\theta\frac{\partial w}{\partial\theta}+(\cot\theta v_\phi-v_z)\left(v_\phi\frac{\partial w}{\partial v_\theta}-v_\theta\frac{\partial w}{\partial v_\phi}\right)-\frac{\mu F}{I}\sin\theta\frac{\partial w}{\partial v_\theta}, \qquad \ldots\ldots(3.8)$$

$$\partial w/\partial\phi=\partial w/\partial\chi=0. \qquad \ldots\ldots(3.9)$$

If we carry out a Fourier transformation of w over the three velocity components v_r and a substitution analogous to (2.12)

$$\Psi(\theta,u_\theta,u_\phi,u_z)=\exp(kT\textstyle\sum u_r^2/2I)\displaystyle\iiint w(\theta,v_\theta,v_\phi,v_z)$$
$$\times\exp(-i\textstyle\sum u_r v_r)\,dv_\theta\,dv_\phi\,dv_z, \qquad \ldots\ldots(3.10)$$

the expression (3.9) transforms into

$$\frac{D'\Psi}{D't}\equiv\frac{\partial\Psi}{\partial t}+i\frac{\partial^2\Psi}{\partial\theta\partial u_\theta}-\frac{iu_\theta}{I}\left[kT\frac{\partial\Psi}{\partial\theta}+\mu F\sin\theta\Psi\right]$$
$$+i\left[\cot\theta\frac{\partial}{\partial u_\phi}-\frac{\partial}{\partial u_z}-\frac{kT}{I}(\cot\theta u_\phi-u_z)\right]\left(u_\phi\frac{\partial\Psi}{\partial u_\theta}-u_\theta\frac{\partial\Psi}{\partial u_\phi}\right). \qquad \ldots\ldots(3.11)$$

The fluctuation terms become correspondingly

$$D'\Psi/D't=-\beta\textstyle\sum u_k\,\partial\Psi/\partial u_k \qquad \ldots\ldots(3.12)$$

for the Brownian motion case (A), with obvious extensions of (2.15) for the other mechanisms. The equations (3.11), (3.12) are again most easily solved by a transformation to polar coordinates in u-space and a separation of variables for Ψ; thus with

$$u_z = u\cos\theta_u,\quad u_\theta = u\sin\theta_u\cos\phi_u,\quad u_\phi = u\sin\theta_u\sin\phi_u \qquad \ldots\ldots(3.13)$$

Ψ can be expressed by

$$\begin{aligned}8\pi^2\Psi = 1 &+ \cos\theta\,\psi_1(u,t) + \sin\theta\sin\theta_u\cos\phi_u\,\psi_2(u,t)\\ &+ (\cos\theta\cos^2\theta_u + \sin\theta\sin\theta_u\cos\theta_u\cos\phi_u)\psi_3(u,t) + \ldots\end{aligned} \qquad \ldots\ldots(3.14)$$

where again the ψ_i are of order $\mu F/kT$ and higher terms are neglected. Elementary, but very lengthy calculations lead to a set of three differential equations with the independent variables u and t. In the Brownian motion case we obtain

$$\left.\begin{aligned}\frac{\partial\psi_1}{\partial t} + i\frac{\partial\psi_2}{\partial u} + i\frac{\psi_2}{u} - i\frac{ukT\psi_2}{I} &= -\beta u\frac{\partial\psi_1}{\partial u},\\ \frac{\partial\psi_2}{\partial t} - i\frac{\partial\psi_1}{\partial u} + i\frac{ukT\psi_1}{I} + i\frac{\psi_3}{u} &= -\beta u\frac{\partial\psi_2}{\partial u} + iu\frac{\mu F}{I},\\ \frac{\partial\psi_3}{\partial t} - i\frac{\partial\psi_2}{\partial u} + i\frac{\psi_2}{u} + i\frac{ukT\psi_2}{I} &= -\beta u\frac{\partial\psi_3}{\partial u}.\end{aligned}\right\} \qquad \ldots\ldots(3.15)$$

For the steady state case $F = F_0 e^{i\omega t}$ the power series development (2.22) and the corresponding expansion

$$\psi_3 = (-\mu F_0/kT)e^{i\omega t}[c_2(\beta u)^2 + c_4(\beta u)^4 + \ldots] \qquad \ldots\ldots(3.16)$$

lead, with the use of (2.25), to the recurrence relations

$$i\omega' a_0 + 2b_1 = i\omega', \qquad \ldots\ldots(3.17a)$$

$$(i\omega' + n)a_n + (n+2)b_{n+1} - \gamma b_{n-1} = 0, \qquad (n>0) \qquad \ldots\ldots(3.17b)$$

$$(i\omega' + n)b_n + (n+1)a_{n+1} - \gamma a_{n-1} - c_{n+1} = 0, \qquad \ldots\ldots(3.17c)$$

$$(i\omega' + n)c_n - nb_{n+1} + \gamma b_{n-1} = 0. \qquad \ldots\ldots(3.17d)$$

By means of (3.17b) and (3.17d) c_n can be expressed in terms of a_n and b_{n-1}; if the corresponding value is substituted in (3.17c), the equation becomes

$$\left[i\omega' + n + \frac{2\gamma}{(n+3)(i\omega'+n+1)}\right]b_n + \left(n+2-\frac{2}{n+3}\right)a_{n+1} - \gamma a_{n-1} = 0. \quad \ldots.(3.18)$$

The system of three-term recurrence relations (3.17a), (3.17b), (3.18) can again be solved by the standard continued fraction method; for the polarization $P^*(\omega)$ the result is, in view of (2.23):

$$\begin{aligned}\frac{P^*(\omega)}{P_0} = 1 - \frac{i\omega'|}{|i\omega'} &+ \frac{2\gamma\quad|}{|1+i\omega'+\frac{1}{2}\gamma/(2+i\omega')} + \frac{(3-\frac{1}{2})\gamma|}{|2+i\omega'}\\ &+ \frac{4\gamma\quad|}{|3+i\omega'+\frac{1}{3}\gamma/(2+i\omega')} + \frac{(5-\frac{1}{3})\gamma|}{|4+i\omega'} + \ldots.\end{aligned} \qquad \ldots\ldots(3.19)$$

For the other collision mechanisms, (3.17) and (3.18) are easily modified by a comparison of (2.24), (2.28) and (2.31); with the abbreviations (2.27) the result becomes in case (B)

$$\begin{aligned}\frac{P^*(\omega)}{P_0} = 1 - \frac{i\omega''|}{|i\omega''} &+ \frac{2\gamma_1\quad|}{|1+i\omega''+\frac{1}{2}\gamma_1/(1+i\omega'')} + \frac{(3-\frac{1}{2})\gamma_1|}{|1+i\omega''}\\ &+ \frac{4\gamma_1\quad|}{|1+i\omega''+\frac{1}{3}\gamma_1/(1+i\omega'')} + \ldots\end{aligned} \qquad \ldots\ldots(3.20)$$

and for the very large collisions (C)

$$\frac{P^*(\omega)}{P_0}=1-\frac{i\omega''|}{|1+i\omega''}+\frac{2\gamma_1\quad|}{|1+i\omega''+\tfrac{1}{2}\gamma_1/(1+i\omega'')}+\ldots. \qquad \ldots\ldots(3.21)$$

§ 4. RESULTS AND DISCUSSIONS

The frequency response of the polarization of a system of rigid dipoles rotating freely in space except for interaction of the Brownian motion type with a viscous substrate is described by (2.25), (2.26) for needle-shaped dipoles, and by (2.25), (3.19) for spherical molecules. These continued fractions are similar to the corresponding expression for the model with free rotation about fixed axes (Gross 1955), which with the abbreviation (2.25) can be written as

$$\frac{P^*(\omega)}{P_0}=1-\frac{i\omega'|}{|i\omega'}+\frac{\gamma\quad|}{|1+i\omega'}+\frac{2\gamma\quad|}{|2+i\omega'}+\frac{3\gamma\quad|}{|3+i\omega'}+\ldots. \qquad \ldots\ldots(4.1)$$

In the limit

$$\gamma\ll 1, \qquad \omega'=\omega/\beta\ll 1 \qquad \ldots\ldots(4.2)$$

the Debye relation for $P^*(\omega)=P'(\omega)-iP''(\omega)$, i.e.

$$\frac{P^*(\omega)}{P_0}=\frac{1}{1+i\omega\tau}\,;\quad \frac{P'(\omega)}{P_0}=\frac{1}{1+\omega^2\tau^2},\quad \frac{P''(\omega)}{P_0}=\frac{\omega\tau}{1+\omega^2\tau^2} \qquad \ldots\ldots(4.3)$$

is obtained in all cases with $\tau=\tau_s$ defined in (2.1) for rotation in space and

$$\tau=\tau_p=I\beta/kT \qquad \ldots\ldots(4.4)$$

for plane rotation. More satisfactory approximations, however, valid for $\gamma\ll 1$ and all values of ω are the second convergents of (2.26), (3.19) and (4.1), which in view of (2.1) and (4.4) can be expressed as

$$\frac{P^*(\omega)}{P_0}=\frac{1}{1+i\omega\tau-\omega^2\tau\beta^{-1}}=\frac{1-\omega^2\tau\beta^{-1}-i\omega\tau}{(1-\omega^2\tau\beta^{-1})^2+\omega^2\tau^2}. \qquad \ldots\ldots(4.5)$$

In the limit of high frequencies this expression becomes proportional to $1/\omega^2$, and its imaginary part (the loss factor) to $1/\omega^3$, in agreement with the general requirements stated in I and by Gross (1955). The result (4.5) with (2.1) or (4.4) also applies to the models with the other collision mechanisms which leave the orientation of the dipoles unaltered, provided the viscosity parameter β is taken to be β_1 in case (B) and $\beta_1(1-\Gamma)$ in case (D). This wide range of validity of the approximation formula (4.5) follows generally from a modification of Smoluchowski's diffusion equation recently proposed by the writer (Sack 1956).

It is of interest to note that Powles (1948) although using an incorrect analysis of the inertia terms, has derived a formula for the polarization, which in the notation of the present paper is equivalent to

$$P^*(\omega)/P_0=[1+i\omega\tau-\omega^2\tau\beta^{-1}+(1-\alpha)i\omega\beta^{-1}]^{-1} \qquad \ldots\ldots(4.6)$$

where α is a parameter lying between zero and unity. The expression (4.5), which has been given in dimensionless units by Gross (1955), results from (4.6) if $\alpha=1$; for $\alpha=0$, Powles' formula becomes

$$\frac{P^*(\omega)}{P_0}=\frac{1}{(1+i\omega\tau)(1+i\omega\beta^{-1})}=\frac{1-\omega^2\tau\beta^{-1}-i\omega(\tau+\beta^{-1})}{(1+\omega^2\tau^2)(1+\omega^2\beta^{-2})}, \qquad \ldots\ldots(4.7)$$

an expression found by Dmitriev and Gurevich (1946) and recently rederived by means of a Lagrangian formalism by Saito and Kato (1956). A result previously given by Rocard (1933) is in fact entirely equivalent to (4.7), but his expression for the real part $P'(\omega)$ as printed contains a plus sign in the numerator.† The differences between the expressions (4.5) and (4.7) are of the order γ, and hence insignificant in most practical cases. In both cases $P^*(\omega)$ changes sign when $\omega^2=\beta/\tau$. In the table the polarizations calculated according to (4.5) and the (corrected Rocard) expression (4.7) are compared with the Debye results (4.3) for the case $\beta\tau=20$, i.e. $\gamma=0{\cdot}05$ for plane rotation or $\gamma=0{\cdot}025$ for the three-dimensional models. A comparative graph for the loss factor $P''(\omega)$ according to the three formulae has been given by Powles (1948) for the case $\beta\tau=10$.

Frequency Dependence of Polarization according to Various Formulae for the case $\beta\tau=20$

$\omega\tau$	$10^3P^*(\omega)/P_0$ Debye	$10^3P^*(\omega)/P_0$ Rocard	$10^3P^*(\omega)/P_0$ eqn (4.5)	10^3C_N Case (A)	10^3C_N Case (B)
0·5	$800-400i$	$789-419i$	$806-408i$	$-8-6i$	$-15-10i$
1	$500-500$	$450-524$	$499-526$	$-13+1$	$-24+2$
2	$200-400$	$158-416$	$172-431$	$-7+8$	$-13+16$
3	$100-300$	$54-308$	$59-323$	$-2{\cdot}2+7{\cdot}5$	$-4+16$
4	$58{\cdot}8-235{\cdot}3$	$11{\cdot}4-239{\cdot}2$	$12{\cdot}5-249{\cdot}4$	$0+6{\cdot}1$	$+0{\cdot}9+12{\cdot}0$
6	$27{\cdot}1-162{\cdot}2$	$-19{\cdot}3-156{\cdot}2$	$-21{\cdot}8-163{\cdot}8$	$+1{\cdot}4+3{\cdot}4$	$4{\cdot}0+6{\cdot}7$
8	$15{\cdot}4-123{\cdot}1$	$-29{\cdot}2-111{\cdot}4$	$-32{\cdot}0-116{\cdot}1$	$1{\cdot}9+2{\cdot}0$	$4{\cdot}2+3{\cdot}3$
10	$9{\cdot}9-99{\cdot}0$	$-31{\cdot}7-83{\cdot}2$	$-34{\cdot}5-86{\cdot}2$	$1{\cdot}7+1{\cdot}1$	$3{\cdot}6+1{\cdot}3$
15	$4{\cdot}4-66{\cdot}4$	$-29{\cdot}0-44{\cdot}6$	$-31{\cdot}0-45{\cdot}4$	$1{\cdot}1+0$	$1{\cdot}8-0{\cdot}5$
20	$2{\cdot}5-49{\cdot}9$	$-23{\cdot}7-26{\cdot}1$	$-25{\cdot}0-26{\cdot}3$	$0{\cdot}5-0{\cdot}2$	$0{\cdot}7-0{\cdot}6$
50	$0{\cdot}4-20{\cdot}0$	$-6{\cdot}8-2{\cdot}9$	$-6{\cdot}9-2{\cdot}8$	0	0
100	$0{\cdot}1-10{\cdot}0$	$-1{\cdot}9-0{\cdot}4$	$-1{\cdot}9-0{\cdot}4$	0	0

The precise nature of the molecular model and of the collision mechanism employed affects the results only in higher approximation. In the last two columns the correction terms C_N (multiplied by 10^3) are given which must be added to (4.5) to give the correct results (2.26) and (2.30) for the rotating needle model with Brownian collisions or with collision mechanism (B) respectively. For small values of γ the correction terms C_P for plane rotation and C_S for the rotating sphere are related to C_N for the same $\omega\tau$ and $\beta\tau$ by

$$C_P\simeq 2C_N;\qquad C_S\simeq \tfrac{3}{2}C_N \qquad \ldots\ldots(4.8)$$

whichever collision mechanism (A), (B) or (D) is operative. It can be seen that for frequencies near the absorption maximum ($\omega\tau\simeq 1$) the correction terms C_N are of the same order as the differences between the Debye values (4.3) and the second approximation (4.5) or (4.7). With increasing frequency the correction terms rapidly decrease, and (4.5) becomes a very good approximation.

With decreasing friction, i.e. with increasing value of γ or γ_1 the differences between the various models become more and more marked but, as pointed out by Gross (1955), in this case quantum effects become important and the classical approach is no longer adequate.

Although the frequency responses (2.26) and (4.1) of the rotating disc and needle models subject to Brownian collisions are similar, the analytic natures of the solutions are essentially different. It was shown in I that the behaviour of the plane rotation model can be expressed by a series of rate constants λ_l which form an arithmetical progression

---

† Professor Rocard (personal communication) has confirmed that this sign has been misprinted.

$$\lambda_l = (\tau_l)^{-1} = \beta(\gamma + l), \qquad \ldots\ldots(2.9)$$

the corresponding amplitudes having alternate signs. It has not been possible to find an analytic solution for the rotating needle model, but the rate constants λ_l are the negatives of the characteristic values of the system of equations (2.18), (2.19(A)), i.e. of those values of $i\omega$ for which the equations (2.24) become compatible when all the right-hand sides vanish. These characteristic values can be expressed as power series in γ by the usual perturbation methods (cf. for example Courant and Hilbert 1953). The results differ according to whether l is even or odd:

$$\left.\begin{aligned} \lambda_{2m} &= \beta[2m + 2\gamma + 2\gamma^2(8m+1) + \ldots]; \\ \lambda_{2m+1} &= \beta[2m + 1 - 4\gamma^2(m+1) - \ldots]. \end{aligned}\right\} \qquad \ldots\ldots(4.10)$$

With increasing γ each even rate constant approaches, and eventually coalesces with, the next higher odd one, after which both constants become complex; the corresponding component of the decay function corresponds to a strongly damped harmonic oscillation; for any finite value of γ there is a finite limit of m, beyond which all the λ_{2m}, λ_{2m+1} are complex. The leading rate constant λ_0, in particular, is found to be

$$\lambda_0 = 1/\tau_0 = \beta(2\gamma + 2\gamma^2 + 16\gamma^3/3 + \ldots) \qquad \ldots\ldots(4.11)$$

and agrees with the reciprocal of the Debye value for the relaxation time τ_S of (2.1) only in first approximation. The critical value γ_c of γ for which λ_0 and λ_1 coalesce has been determined numerically to $\gamma_c = 0{\cdot}19$ with $\lambda_0 = 0{\cdot}66\beta$; for larger values of γ the overall response of the system resembles that of a strongly damped harmonic oscillator. This must be interpreted as meaning that in view of the large inertial effects a displaced distribution will not only tend towards equilibrium, but overshoot it. For the plane rotation model the approach to equilibrium is always asymptotic, however large γ. Since, however, for polar liquids and solutions γ is several orders of magnitude smaller than γ_c, this mathematical effect is of no practical importance.

In case (B), which corresponds to the non-polar molecules being considerably heavier than the dipoles, the analytical solution (2.35) with (2.34) shows great similarity to the corresponding plane rotation model; as in equations (3.27), (3.25) of I there is only one real value of $-\lambda = i\omega$ for which the denominator of (2.35) vanishes; for large values of X, i.e. small values of γ_1, the asymptotic expansion of the exponential integral (cf. Magnus and Oberhettinger 1949)

$$e^X[-\mathrm{Ei}(-X)] = \frac{1}{X} - \frac{1}{X^2} + \frac{2\,!}{X^3} - \ldots \qquad \ldots\ldots(4.12)$$

leads with (2.34) to

$$\lambda = 1/\tau = \beta_1(2\gamma_1 + 8\gamma_1^3 + \ldots). \qquad \ldots\ldots(4.13)$$

As in the rotating disc model the relaxation times calculated for the Brownian motion case from (4.11) and for collision mechanism (B) from (4.13) agree only to the first power in γ or γ_1.

It remains to discuss briefly the results for case (C), in which the dipoles can change their orientation abruptly. The continued fraction expansions (2.32) and (3.21) show the same trend as the corresponding expression for the rotating disc (Gross 1955); the first approximation gives a Debye behaviour with $\tau=\beta_1^{-1}$ in all cases, and the deviations therefrom increase with γ_1. For the dielectric after-effect $P_a(t)$ the formula (A10) of the Appendix with the definition (A7) shows that initially the polarization decays exponentially with a rate constant β_1, but since for large z

$$ {}_1F_1(1\,;\,\tfrac{1}{2}\,;\,-z^2) \sim -1/(2z^2) \qquad \ldots\ldots(4.14)$$

(cf. Magnus and Oberhettinger 1949), it eventually changes sign and decreases as

$$\lim P_a(t)/P_0 = -(I/kTt^2)\exp(-\beta_1 t). \qquad \ldots\ldots(4.15)$$

If γ_1 is small, this change of sign will occur when most of the polarization has already decayed, and the decay can be described as essentially exponential.

ACKNOWLEDGMENT

This work forms part of the fundamental research programme of the British Rayon Research Association.

APPENDIX

Derivation of (2.33)

An analytic expression for the polarization of the rotating needle model under the influence of the very large collisions (C) is most easily derived for the dielectric after-effect $P_a(t)$, for which $F=F_0$ $(t<0)$, $F=0$ $(t>0)$. With the substitution

$$\psi_r=(\mu F_0/kT)\exp(-\beta_1 t+kTu^2/2I)y_r \qquad (r=1,\,2) \qquad \ldots\ldots(\mathrm{A1})$$

(2.18) and (2.19(C)) yield for $\tau>0$

$$\frac{\partial y_1}{\partial t}+i\left(\frac{\partial y_2}{\partial u}+\frac{y_2}{u}\right)=0, \qquad \frac{\partial y_2}{\partial t}-i\frac{\partial y_1}{\partial t}=0 \qquad \ldots\ldots(\mathrm{A2})$$

with the initial conditions

$$y_1(u,0)=\exp(-kTu^2/2I); \qquad y_2(u,0)=0. \qquad \ldots\ldots(\mathrm{A3})$$

On elimination of y_2, (A2) becomes

$$\frac{\partial^2 y_1}{\partial t^2}-\frac{\partial^2 y_1}{\partial u^2}-u^{-1}\frac{\partial y}{\partial u}=0 \qquad \ldots\ldots(\mathrm{A4})$$

of which the general solution which is analytic at $u=0$ has been given by Laplace and can be written

$$y_1(u,t)=y_1(0,t)+\frac{u^2}{2^2}\frac{\partial^2 y_1(0,t)}{\partial t^2}+\frac{u^4}{2^2 4^2}\frac{\partial^4 y_1(0,t)}{\partial t^4}+\ldots \qquad \ldots\ldots(\mathrm{A5})$$

(cf. Bateman 1944). For $t=0$ this series must be equal to the expansion of (A3); this yields

$$y_1(0,t)=1-2z^2+\frac{2^2 z^4}{1.3}-\frac{2^3 z^6}{1.3.5}+\ldots={}_1F_1(1\,;\,\tfrac{1}{2}\,;\,-z^2) \qquad \ldots\ldots(\mathrm{A6})$$

where

$$z^2=kTt^2/2I \qquad \ldots\ldots(\mathrm{A7})$$

and ${}_1F_1$ denotes the confluent hypergeometric series. If Kummer's transformation is applied to (A6) (Magnus and Oberhettinger 1949), the result involves the error function of imaginary argument

$$y_1(0,t)=\exp(-z^2)\,{}_1F_1(-\tfrac{1}{2};\tfrac{1}{2};z^2)=\exp(-z^2)\left(1-z^2-\frac{z^4}{3.2!}-\frac{z^6}{5.3!}\cdots\right) \qquad \text{......(A8)}$$

$$=1-2z\exp(-z^2)\int_0^z \exp(s^2)\,ds=1-\pi^{1/2}z\exp(-z^2)[i\,\mathrm{Erf}(-iz)]. \qquad \text{......(A9)}$$

This, with (2.20) and (A1), gives the result for the polarization $P_a(t)$

$$\frac{P_a(t)}{P_0}=\exp(-\beta_1 t)\{1-\pi^{1/2}z\exp(-z^2)[i\,\mathrm{Erf}(-iz)]\}. \qquad \text{......(A10)}$$

The steady state response to an alternating field is

$$P^*(\omega)=P_0-i\omega\int_0^\infty P_a(t)e^{-i\omega t}\,dt. \qquad \text{......(A11)}$$

From tables of Laplace transforms (Erdélyi 1954, p. 176) we find

$$L[\exp(-\alpha^2t^2)\,\mathrm{Erf}(i\alpha t)]=(2\alpha i\pi^{1/2})^{-1}\exp\left(\frac{p^2}{4\alpha^2}\right)\mathrm{Ei}\left(-\frac{p^2}{4\alpha^2}\right); \qquad \text{......(A12)}$$

this expression and the general rules for Laplace transforms (Erdélyi 1954, p. 129) applied to (A7), (A10) and (A11) lead to (2.33), (2.34).

REFERENCES

BATEMAN, H., 1944, *Partial Differential Equations of Mathematical Physics* (New York : Dover Publications).

CHANDRASEKHAR, S., 1943, *Rev. Mod. Phys.*, **15**, 1.

COURANT, R., and HILBERT, D., 1953, *Methods of Mathematical Physics*, Vol. I, Ch. I, (New York : Interscience Publishers).

DEBYE, P., 1945, *Polar Molecules* (New York : Dover Publications).

DMITRIEV, V. A., and GUREVICH, S. B., 1946, *Zhur. Eksp. Teor. Fiz.*, **16**, 937.

ERDÉLYI, A. (Ed.), 1954, *Tables of Integral Transforms*, Vol. I (Bateman Manuscript Project; New York : McGraw-Hill).

GREEN, M. S., 1952, *J. Chem. Phys.*, **20**, 1281.

GROSS, E. P., 1955, *J. Chem. Phys.*, **23**, 1415.

KEILSON, J., and STORER, J. E., 1952, *Quart. Appl. Math.*, **10**, 243.

MAGNUS, H., and OBERHETTINGER, F., 1949, *Formulas and Theorems of the Special Functions of Mathematical Physics* (New York : Chelsea Publishing Co.).

PERRON, O., 1950, *Die Lehre von den Kettenbrüchen* (New York : Chelsea Publishing Co.).

POWLES, J. G., 1948, *Trans. Faraday Soc.*, **44**, 802, 844.

ROCARD, M. Y., 1933, *J. Phys. Radium*, **4**, 247.

SACK, R. A., 1953, *Kolloid Z.*, **134**, 83; 1956, *Physica*, **22**, 917; 1957, *Proc. Phys. Soc.* B, **70**, 414.

SAITO, N., and KATO, T., 1956, *Busseiron Kenkyu*, **92**, 103.

SLATER, J. C., and FRANK, N. H., 1947, *Mechanics*, Chapter VI (New York : McGraw-Hill).

RESEARCH NOTES

Examples of the Maxwell–Wagner Type of Dielectric Absorption using Wool Wax–Water Mixtures

By J. S. DRYDEN and R. J. MEAKINS
Division of Electrotechnology, National Standards Laboratory,
Commonwealth Scientific and Industrial Research Organisation

MS. received 6th February 1956, *and in revised form* 12*th October* 1956

THE dielectric properties of a system in which spherical regions of conducting impurity are distributed in a perfect insulating material were examined theoretically by Wagner (1914). This work dealt with the special case of low concentration of the conducting phase. The first attempt at an experimental investigation of a solid dielectric conforming to Wagner's model was made by Sillars (1937) who examined the properties of a dispersion of water in paraffin wax. Sillars did not succeed in distributing the water as spheres and the dielectric absorption in his mixtures did not agree with theoretical predictions.

A successful synthesis of the Wagner model was made by Hamon (1953), who distributed 3% of approximately spherical particles of an organic semiconductor (copper phthalocyanine) in paraffin wax and obtained reasonably good agreement with Wagner's theory.

More recently, Kharadly and Jackson (1953) have extended Wagner's theory to allow for interaction between the spheres of conducting material in the special case of a cubic array. They measured the dielectric absorption in a model, made on a macroscopic scale, consisting of spheres of a conducting polystyrene–nitrobenzene mixture in a continuous phase of polythene. Good agreement with theory was obtained for concentrations up to 30%.

The work to be described in this note arose from an investigation of the dielectric properties of wool wax. The latter is a complex mixture of organic esters (Truter 1951) obtained from the scouring of greasy wool, and having a colour and consistency similar to unrefined petroleum jelly. For the purpose of the present work, the important property of wool wax is its ability to form water-in-oil emulsions. It was found that in thoroughly mixed samples of water and wool wax the water was distributed in the form of spheres and a high concentration of water in the wax could be obtained. It therefore seemed likely that such mixtures would provide an example of the dielectric absorption in mixtures containing spherical regions of conducting impurity.

The wool wax used was commercial material designated Adeps Lanae (anhydrous grade). This was purified by washing twice with distilled water whilst being agitated with steam. The water was separated off and the wax dried by heating at 110°c under a pressure of 1 mm of mercury. The petroleum jelly used in the later experiments was high-grade, amber-coloured, commercial material.

The mixtures were prepared by stirring the wax and water vigorously in a beaker with a small wire spatula.

The dielectric loss factors of mixtures of 0, 10, 20, 30, 40% by weight (0, 9·3, 18·7, 28·3, 38·1% by volume) of water in the wool wax are shown in figure 1.

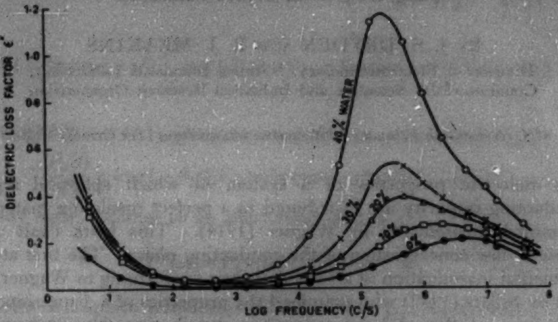


Figure 1. Dielectric absorption in mixtures of wool wax and water, measured at 20°c.

Because of the magnitude of the dielectric loss in the wool wax itself and the relatively small effect of low concentrations of water, experiments were not conducted on mixtures with below 10% water content. At the concentrations used, the expressions obtained by Wagner will not apply because of interaction between the conducting regions, and the results are considered in relation to the expression given by Kharadly and Jackson (1953) for the complex permittivity $\epsilon = \epsilon' - j\epsilon''$ of the mixture:

$$\frac{\epsilon}{\epsilon_1} = 1 + \frac{[(\epsilon_2 - \epsilon_1)/(\epsilon_2 + 2\epsilon_1)]3v}{1 - [(\epsilon_2 - \epsilon_1)/(\epsilon_2 + 2\epsilon_1)]v}, \qquad \ldots\ldots(1)$$

where ϵ_1 and ϵ_2 are the complex permittivities of the continuous phase and of the conducting material respectively and v the volume fraction of conducting material. This equation has been derived for the special case of a cubic array of dipoles and taking into account interaction between the dipoles induced in the individual elements.

Substituting $\epsilon_2' - j4\pi\sigma_2/\omega$ for ϵ_2 in the above expression where σ_2 is the conductivity (in c.g.s. units) of the dispersed material, and assuming no dielectric absorption in the continuous phase (i.e. $\epsilon_1 = \epsilon_1'$), the following expression can be deduced for the dielectric loss factor ϵ'' of the mixture:

$$\epsilon'' = \epsilon_1' \frac{9v\gamma \dfrac{\omega\tau}{1+\omega^2\tau^2}}{1 - 2v\dfrac{1+\beta\omega^2\tau^2}{1+\omega^2\tau^2} + v^2\left\{\left(\dfrac{1+\beta\omega^2\tau^2}{1+\omega^2\tau^2}\right)^2 + 9\dfrac{\gamma^2\omega^2\tau^2}{1+\omega^2\tau^2}\right\}}, \qquad \ldots\ldots(2)$$

where ω is the angular frequency, τ $(=(2\epsilon_1' + \epsilon_2')/4\pi\sigma_2)$ is the relaxation time for low concentration of impurity, β is $(\epsilon_2' - \epsilon_1')/(\epsilon_2' + 2\epsilon_1')$ and γ is $3\epsilon_1'/(\epsilon_2' + 2\epsilon_1')$.

The dielectric loss factor passes through a maximum given, with sufficient accuracy, by the following expression:

$$f_{\max} = \frac{2\sigma_2}{2\epsilon_1' + \epsilon_2'}\left\{1 - v\frac{3\epsilon_1'}{\epsilon_2' + 2\epsilon_1'}\right\}. \qquad \ldots\ldots(3)$$

Two comparisons can be made with theory: firstly, the frequency of maximum absorption and, secondly, the magnitude of the absorption.

Substituting for σ_2 the measured value of the conductivity of the distilled water, $1\cdot2\times10^{-6}$ ohm^{-1} cm^{-1} ($1\cdot1\times10^{6}$ c.g.s. units), and taking 80 for ϵ_2' and 3 for ϵ_1', $f_{\max}$ becomes 24 kc/s. The frequency of maximum absorption obtained in these measurements, as estimated from the higher water contents, is greater than this by a factor of approximately 11. It seemed likely that the discrepancy was due to the presence of ionic matter introduced into the water during the mixing procedure. This was confirmed by the higher frequencies of maximum absorption obtained with crude wool wax and also by the increased conductivity found in samples of water separated from mixtures with wool wax.

In the table, the magnitude of the dispersion $\epsilon_s'-\epsilon_\infty'$ is compared with the value calculated from equation (1), for the different water contents. ϵ_s' is the permittivity below the absorption region and ϵ_∞' that above the absorption region. The values of ϵ_s' listed in the table are those measured at low frequencies and $\epsilon_s'-\epsilon_\infty'$ was obtained, in each case, by integrating the area under the curve, since

$$\epsilon_s'-\epsilon_\infty'=\frac{2}{\pi}\int\epsilon''(f)d(\log f). \qquad \ldots\ldots(4)$$

The calculated values of ϵ_s' and ϵ_∞' were obtained from the following expressions which can be deduced from equation (1) in the special cases when $\omega\to0$ and $\omega\to\infty$ respectively:

$$\epsilon_s'=\epsilon_1'\left\{1+\frac{3v}{1-v}\right\}, \qquad \ldots\ldots(5a)$$

$$\epsilon_\infty'=\epsilon_1'\left\{1+\frac{[(\epsilon_2'-\epsilon_1')/(\epsilon_2'+2\epsilon_1')]3v}{1-[(\epsilon_2'-\epsilon_1')/(\epsilon_2'+2\epsilon_1')]v}\right\}. \qquad \ldots\ldots(5b)$$

A different value must be used for ϵ_1' in equation (5*a*) from that in equation (5*b*), because of the dispersion in the wool wax: in this way, the absorption in the continuous phase is taken into account. The value of ϵ_1' to be used in equation (5*a*) was that measured at low frequencies but, because of the small dispersion and the fact that it was difficult to measure relative permittivity with sufficient accuracy above the absorption region, the ϵ_1' to be used in equation (5*b*) was arrived at by subtracting the dispersion in the wool wax obtained by the use of equation (4) from 3·2, the value measured at low frequencies. Reference to the table shows that the agreement between experiment and equation (1) is good for up to 30% water content.

(1)	(2)	(3)		(4)	
		(*a*)	(*b*)	(*a*)	(*b*)
10	0·093	4·2	4·3	1·13	1·08
20	0·187	5·2	5·45	1·56	1·46
30	0·283	6·9	7·0	2·03	2·05
40	0·381	11·1	9·1	3·82	2·89

(1) % water by weight; (2) volume fraction v; (3) ϵ_s', (4) ϵ_s'-ϵ_∞': (*a*) measured (*b*) calculated from equation (5).

The magnitude of the dielectric absorption in the wool wax prevents the mixtures discussed in the previous paragraphs from being ideal examples of the dielectric absorption to be expected from the Wagner model. In an attempt to overcome this, it was decided to use as the continuous phase a mixture of petroleum jelly and sufficient wool wax to enable an emulsion to be formed. Figure 2 gives the results obtained with various water concentrations in a mixture of 3% wool wax and petroleum jelly. With this low concentration of wool wax it was only possible to obtain emulsions with up to about 20% water. In further experiments, which gave similar results, petroleum jelly containing 10% of wool wax was used, and this enabled up to 40% of water to be taken up.

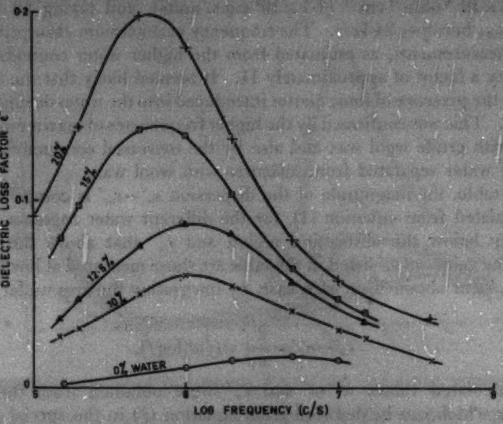


Figure 2. 3% wool wax in petroleum jelly. Effect of adding various proportions of water. Measurement at 20°c.

In contrast to the case where water was dispersed in undiluted wool wax, the area under the absorption curve in these mixtures is higher than that calculated from equation (1): for example, at 10% water content the absorption is approximately 25% higher, and for 20% water content the measured absorption is approximately 70% higher, than that calculated. Examination under the microscope showed that for the mixtures in petroleum jelly, in contrast to those in wool wax where the particles were all about 0·5 μ in diameter, there was a large variation in the diameter of the spheres (0·5 to 5·5 μ). It can be concluded then that, provided the diameters of the conducting particles have a narrow distribution, the dielectric properties calculated on the basis of a cubic array are a good approximation up to a concentration of about 30%, but are not so when there is a wide distribution. Pearce (1955), in experiments on an electrolytic tank analogue of a random distribution of spherical conductors, found that the conductivity, for the same concentration, depended markedly on the distribution in the diameters of the spheres.

REFERENCES

HAMON, B. V., 1953, *Aust. J. Phys.*, **6**, 304.
KHARADLY, M. M. Z., and JACKSON, W., 1953, *Proc. Instn Elect. Engrs*, **100** (III), 199.
PEARCE, C. A. R., 1955, *Brit. J. Appl. Phys.*, **6**, 113.
SILLARS, R. W., 1937, *J. Instn Elect. Engrs*, **80**, 378.
TRUTER, E. V., 1951, *Quart. Rev. Chem. Soc., Lond.*, **5**, 390.
WAGNER, K. W., 1914, *Arch. Elecktrotech.*, **2**, 371.

A General Theory of the Surface Friction of Solids

By E. H. B. WILLIAMS

Research Laboratories, Joseph Lucas (Electrical) Ltd., Birmingham

MS. received 30th January 1957

In a recent paper under the above title by Rubenstein (1956), a very restricted model for the surface asperities is used in deriving a roughness factor. The purpose of the present note is to show that this is unnecessary. The notation employed is that of the original paper, to which also the numbering of equations refers.

Equation (5), relating the frictional force F and the load L at the contact between plane surfaces, may be written

$$F = S_0 \tfrac{1}{4}\pi(n\Delta^2)^{1-\alpha_1}(L/K_1)^{\alpha_1}.$$

The term $(n\Delta^2)^{1-\alpha_1}$ has a direct and clear physical significance; it expresses the fact that when the deformation is not plastic the real area of contact is greater (other things being equal) the greater the number of asperities over which the load is distributed, and the greater the radius of curvature of the individual asperities. The height and distribution of the asperities are not involved; their introduction in equation (6) gives to the roughness factor a form which obscures its real significance, and also requires the otherwise unnecessary assumption of a geometrically regular surface. The consequent loss of generality is indicated in figure 1.

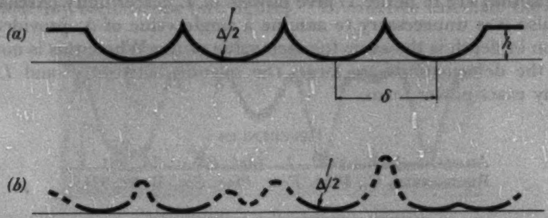


Figure 1. (*a*) Rubenstein's model, (*b*) more general model.

Similar considerations apply to the other expressions given by Rubenstein. Equation (10), for instance, may be written

$$F = \text{const.}\ S_0 D^2(n_0\Delta^2)^{1-\alpha_2}(L/D^2K_1)^{\beta},$$

the constant being purely numerical.

Further, it is not essential to the argument to make the simplifying assumption that the asperities have all the same radius of curvature, provided that the load is small enough for the approximation $d^2/4 \simeq h\Delta$ to be adequate for all values of Δ. When this is the case the power of L in the relations quoted is unchanged by postulating a range of values of Δ.

This statement may be proved as follows, for the fundamental case of contact between a hard smooth plane surface and a plane deformable surface similar to

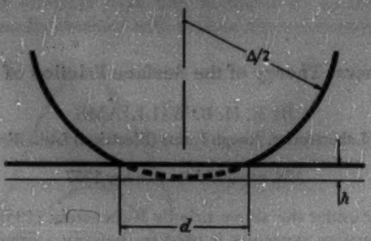


Figure 2.

that of figure 1 (*b*), but with asperities of differing radii. When the hard surface moves through a distance h as shown in figure 2, the load carried by any one asperity is, cf. equation (2),

$$L_i = K_1(d_i^{\alpha_n+2}/\Delta_i^{\alpha_n}) \simeq K_1(4h\Delta_i)^{1/\alpha_n}/\Delta_i^{\alpha_n}.$$

Thus the total load carried is

$$L \simeq K_1(4h)^{1/\alpha_n} \sum \Delta_i^{1-\alpha_n/2}.$$

Also the total frictional force is

$$F = S_0 \sum A_i \simeq S_0 \pi h \sum \Delta_i.$$

Eliminating h between these two equations, we obtain

$$F \simeq S_0 \tfrac{1}{4}\pi[\sum \Delta_i/(\sum \Delta_i^{1-\alpha_n/2})^{\alpha_n}](L/K_1)^{\alpha_n}.$$

The expression in square brackets may be regarded as defining the mean value of $(n\Delta^2)^{1-\alpha_n}$.

Models in which the asperities undergoing deformation are not confined to one level (as they are in figure 1) give powers of L nearer unity (Archard 1953); but here also it is unnecessary to assume a single value of Δ, provided that the distribution in depth is the same for each value of Δ. Where this is not the case, or where the deformations are large, the relation between F and L may not possess any exact power form.

References

Archard, J. F., 1953, *J. Appl. Phys.*, **24**, 981.
Rubenstein, C., 1956, *Proc. Phys. Soc.* B, **69**, 921.

Note on Production of Strong Magnetic Fields of Short Duration and Measurement of their Intensity

By A. PIEKARA, J. MAŁECKI, M. SURMA and J. GIBALEWICZ
Institute of Physics, Polish Academy of Science, Adam Mickiewicz University, Poznan

MS. received 4th December 1956

Transient magnetic fields of great intensity were produced by the present authors by discharging a battery of condensers through a coil. This method of obtaining transient magnetic fields was first discussed by Kapitza (1924) and applied by Wall (1926), Raoult (1949), Myers (1953), Ince (1955) and others.

Lately very interesting results using this method have been obtained by Fürth and Waniek (1956).

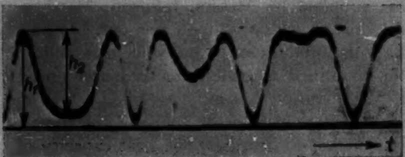
In the authors' device the condenser battery of a capacity of 2 μF and a working voltage of 50 kv (2500 joules) was discharged through a coil connected in series with a special gas-discharge tube (Piekara and Małecki 1957). Some dozens of various multilayer coils were tested before the type yielding the best results could be found. The coils were wound with 100–200 turns of copper wire about 0·7 to 1·1 mm in diameter. The layers were insulated with celluloid sheets 0·5 to 1·0 mm thick. As the inner layer usually underwent mechanical damage, it proved of advantage to wind that layer with somewhat thicker wire.

Two independent methods, using the induction effect and the Faraday effect, were employed for measuring the magnetic field intensity. In the first method the voltage from a search coil of a given effective area S was recorded using a cathode ray oscilloscope. The field intensity was calculated from the relation

$$H = \frac{10^8 F(k)}{S\omega}(E_1 + E_2) \text{ oersteds}$$

where $F(k)$ is a function of the damping coefficient, E_1 and E_2 denote respectively the magnitudes of the first and second extrema of voltage induced in the search coil, and ω is the angular frequency.

The second method (Małecki, Surma and Gibalewicz 1957) was based on Faraday's magneto-optical effect. The cell containing a liquid of known Verdet constant V was placed within the coil between crossed polarizing prisms. The effective optical length l of the cell was calculated using the topography of the field. The magnetic rotation of the plane of polarization caused variations in the intensity of the transmitted light and these were recorded with the use of a photo-multiplier and a cathode-ray oscilloscope (see figure). In order to calculate the maximum field intensity H, the ordinate h_1 corresponding to the rotation angle of $\frac{1}{2}\pi(2n-1)$ ($n = 1, 2, 3 \ldots$) was measured on the oscillograms as well as the ordinate h_2, corresponding to the maximum angle of rotation $\alpha_{\max}$.
Then



A typical oscillogram showing the variation of light intensity. The maximum angle of rotation amounts to 160°.

$$H_{\max} = \frac{1}{Vl}\left[\tfrac{1}{2}n\pi + \arcsin\left(\frac{h_2}{h_1}\right)^{1/2}\right].$$

Rotation angles up to 700° for CS_2 were obtained. Small rotation angles (less than $\frac{1}{2}\pi$) were measured employing an auxiliary biasing rotation of the analyser so that extinction of the light beam was obtained. Results yielded by both methods are in agreement within the limit of error (about 1%).

The maximum field intensity of 350 000 Oe in a volume of about 1 cm^3 was obtained using a five-layer coil having 154 turns mounted in a special well-fitting steel device to avoid damage to the coil (see Piekara and Małecki 1957).

REFERENCES

FÜRTH, H. P., and WANIEK, R. W., 1956, *Rev. Sci. Instrum.*, **27**, 195.
INCE, A. N., 1955, *Proc. Instn Elect. Engrs*, Pt. C, **102**, 25.
KAPITZA, P., 1924, *Proc. Roy. Soc.* A, **105**, 691.
MAŁECKI, J., SURMA, M., and GIBALEWICZ, J., 1957, *Acta Physica Polonica*, in the press.
MYERS, W. R., 1953, *J. Sci. Instrum.*, **30**, 237.
PIEKARA, A., and MAŁECKI, J., 1957, *Acta Physica Polonica*, in the press.
RAOULT, G., 1949, *Ann. Phys., Paris*, **4**, 369.
WALL, T. F., 1926, *J. Instn Elect. Engrs*, **64**, 409.

The Atomic Heat of Samarium from 2° to 20°K

BY LOIS M. ROBERTS

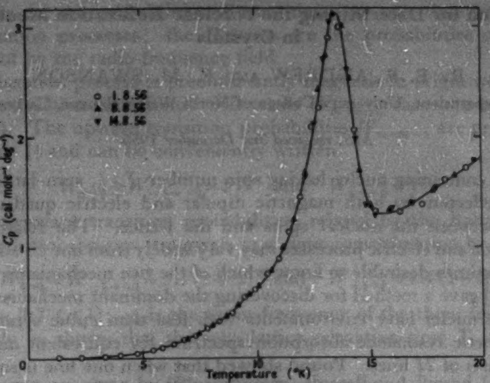
Radar Research Establishment, Malvern, Worcs.

MS. received 10*th December* 1956

RECENT measurements by J. M. Lock (to be published) of the magnetic susceptibility of samarium have shown a pronounced maximum at 14°K, which is believed to indicate the onset of antiferromagnetism, and which should therefore be accompanied by an anomaly in the specific heat. Measurements to investigate this have been made in the temperature range 2° to 20°K.

The samarium—supplied by Messrs. Johnson, Matthey and Co.—was the same sample from which Lock's specimen was cut. It had been prepared by evaporation on to a thin tantalum strip, and since this could not be removed the heat capacity of the combined specimen was measured and a correction for the tantalum applied. The complete sample weighed 6·6 g, and the quantity of tantalum present was estimated from its area and thickness to be 0·8 g, leaving 5·8 g for the weight of the samarium. The heat capacity of the tantalum was calculated using the measurements of Keesom and Desirant (1941) up to 4·2°K and extrapolating their formula ($c_n = 464{\cdot}5(T/246{\cdot}5)^3 + 0{\cdot}00141\,T$ cal mole^{-1} deg^{-1}) to 20°K.

The results (see figure) show a very sharp maximum in the atomic heat at 13·6°K, which corresponds well with the observed maximum in the susceptibility. A smooth curve has been drawn through the experimental points and values read from it are given in the table. At 9·5°K there is a small irregularity which is just outside experimental error. The scatter of the points about the smooth curve is about ±2%, but there may be a bias of up to ±5%. This arises partly in the correction for the calorimeter which, owing to the small quantity of samarium available, had a heat capacity approximately equal to that of the specimen, and partly in the uncertainty in the weight of the samarium owing to the presence of the tantalum. Errors in estimating the heat capacity of the tantalum are negligible.



The atomic heat of samarium

Atomic Heat of Samarium

T	C_p	*T*	C_p	*T*	C_p
2	0·0075	9	0·43	15	1·35
3	0·0136	10	0·63	16	1·31
4	0·026	11	0·90	17	1·40
5	0·052	12	1·46	18	1·53
6	0·104	13	2·55	19	1·67
7	0·179	13·6	3·12	20	1·81
8	0·29	14	2·80		

T in °K, C_p in cal mole^{-1} deg.$^{-1}$.

It would be of interest to know what is the entropy associated with the 13·6°K anomaly, but to obtain this some assumption must be made as to the lattice specific heat. As a first approximation the atomic heat of lanthanum (Parkinson *et al.* 1951) has been subtracted from that of the samarium, and the entropy under the difference curve up to 20°K has been calculated. The value obtained is 0·67 cal mole^{-1} deg^{-1}. An alternative approximation is to use the atomic heat of gadolinium calculated assuming a Debye θ of 152°K (Griffel *et al.* 1954). This gives a value of 0·93 cal mole^{-1} deg^{-1}. It seems probable that the true value is somewhere between these two.

ACKNOWLEDGMENTS

The author wishes to thank Mr. D. C. Handley for his very able experimental assistance, also Dr. J. M. Lock and Dr. D. H. Parkinson for their interest and suggestions. This note is published by permission of the Controller, H.M. Stationery Office.

REFERENCES

GRIFFEL, M., SKOCHDOPOLE, R. E., and SPEDDING, F. H., 1954, *Phys. Rev.*, **93**, 657.

KEESOM, W. H., and DESIRANT, M., 1941, *Physica*, **8**, 273; Commun. Kamerlingh Onnes Lab. Leiden, No. 257 b.

PARKINSON, D. H., SIMON, F. E., and SPEDDING, F. H., 1951, *Proc. Roy. Soc.* A, **207**, 137.

A Method for Determining the Nuclear Relaxation Mechanism in Crystals

By E. R. ANDREW and K. M. SWANSON
Physics Department, University College of North Wales, Bangor, Caernarvonshire

MS. received 3rd December 1956

In crystals containing nuclei having spin number $I > \frac{1}{2}$, spin–lattice relaxation may be effected by both magnetic dipolar and electric quadrupolar interactions between the nuclear spins and the lattice. The relative strengths of the magnetic and electric processes may vary widely from one crystal to another, and it is sometimes desirable to know which of the two mechanisms is dominant. Pound (1950) gave a method for discovering the dominant mechanism in crystals in which the nuclei have environments with less than cubic symmetry. The nuclear magnetic resonance absorption spectrum for equivalent nuclei in such crystals consists of $2I$ lines. Pound showed that when one line is saturated with a strong radio-frequency power applied to a coil wound round the crystal, the strengths of the other lines, measured at low power in a second coil wound orthogonally on the crystal, are markedly affected when the dominant relaxation mechanism is quadrupolar, but are not affected when the dominant mechanism is magnetic. This difference in behaviour stems from transitions for which the magnetic quantum number m changes by 2, which are allowed for quadrupolar interactions but not for magnetic interactions, thus causing the magnetic sub-level populations to be much more interrelated in the quadrupole case. Pound's two-coil method of distinguishing between the two mechanisms is not always convenient and the present note shows that the distinction can also be made from a study of the saturation behaviour of the lines using one coil and one radio frequency. The distinction is brought out particularly well by studying the saturation behaviour when the crystal is so oriented that two lines are superposed.

We begin with an analysis which is similar to Pound's but more general since we shall at first suppose that power is supplied at the frequencies of all the spectral lines. Later we can equate to zero the power applied to any particular lines. Although the argument can be extended to other values of I we treat here the case of $I = \frac{3}{2}$ since the results of the analysis have been applied to the ^{23}Na resonance in a crystal of sodium nitrate. There are four magnetic sub-levels; let the populations of the levels be N_m. Then the sub-level populations obey the following differential equations:

$$\left.\begin{aligned}
\frac{dN_{-\frac{3}{2}}}{dt} &= -N_{-\frac{3}{2}}(W_{-\frac{3}{2}\to-\frac{1}{2}} + W_{-\frac{3}{2}\to\frac{1}{2}} + P_{-\frac{3}{2}\to-\frac{1}{2}}) \\
&\qquad + N_{-\frac{1}{2}}(W_{-\frac{1}{2}\to-\frac{3}{2}} + P_{-\frac{1}{2}\to-\frac{3}{2}}) + N_{\frac{1}{2}}W_{\frac{1}{2}\to-\frac{3}{2}}, \\
\frac{dN_{-\frac{1}{2}}}{dt} &= N_{-\frac{3}{2}}(W_{-\frac{3}{2}\to-\frac{1}{2}} + P_{-\frac{3}{2}\to-\frac{1}{2}}) - N_{-\frac{1}{2}}(W_{-\frac{1}{2}\to-\frac{3}{2}} + W_{-\frac{1}{2}\to\frac{1}{2}} + W_{-\frac{1}{2}\to\frac{3}{2}} \\
&\qquad + P_{-\frac{1}{2}\to-\frac{3}{2}} + P_{-\frac{1}{2}\to\frac{1}{2}}) + N_{\frac{1}{2}}(W_{\frac{1}{2}\to-\frac{1}{2}} + P_{\frac{1}{2}\to-\frac{1}{2}}) + N_{\frac{3}{2}}W_{\frac{3}{2}\to-\frac{1}{2}}, \\
\frac{dN_{\frac{1}{2}}}{dt} &= N_{-\frac{3}{2}}W_{-\frac{3}{2}\to\frac{1}{2}} + N_{-\frac{1}{2}}(W_{-\frac{1}{2}\to\frac{1}{2}} + P_{-\frac{1}{2}\to\frac{1}{2}}) - N_{\frac{1}{2}}(W_{\frac{1}{2}\to-\frac{1}{2}} + W_{\frac{1}{2}\to-\frac{3}{2}} \\
&\qquad + W_{\frac{1}{2}\to\frac{3}{2}} + P_{\frac{1}{2}\to-\frac{1}{2}} + P_{\frac{1}{2}\to\frac{3}{2}}) + N_{\frac{3}{2}}(W_{\frac{3}{2}\to\frac{1}{2}} + P_{\frac{3}{2}\to\frac{1}{2}}), \\
\frac{dN_{\frac{3}{2}}}{dt} &= N_{-\frac{1}{2}}W_{-\frac{1}{2}\to\frac{3}{2}} + N_{\frac{1}{2}}(W_{\frac{1}{2}\to\frac{3}{2}} + P_{\frac{1}{2}\to\frac{3}{2}}) - N_{\frac{3}{2}}(W_{\frac{3}{2}\to-\frac{1}{2}} + W_{\frac{3}{2}\to\frac{1}{2}} + P_{\frac{3}{2}\to\frac{1}{2}}).
\end{aligned}\right\} \quad \ldots\ldots(1)$$

The W terms are the probabilities of transitions per unit time brought about by the relaxation processes; the P terms are the probabilities of transitions brought about by the radio-frequency field.

If the relaxation process is predominantly magnetic in origin we can neglect quadrupole relaxation transitions. All terms $W_{m\to m\pm 2}$ can then be omitted in equations (1). The upward transition probabilities $W_{m\to m-1}$ are proportional to $(I+m)(I-m+1)$ and can be conveniently written

$$W_{\frac{3}{2}\to\frac{1}{2}} = W_{-\frac{1}{2}\to-\frac{3}{2}} = 3W \quad \text{and} \quad W_{\frac{1}{2}\to-\frac{1}{2}} = 4W \qquad \ldots\ldots(2)$$

while the downward transition probabilities, related by the Boltzmann factor to the upward probabilities, are given to a close approximation by

$$W_{\frac{1}{2}\to\frac{3}{2}} = W_{-\frac{3}{2}\to-\frac{1}{2}} = 3W(1+\Delta) \quad \text{and} \quad W_{-\frac{1}{2}\to\frac{1}{2}} = 4W(1+\Delta), \ldots\ldots(3)$$

where $\Delta = h\nu_0/kT$ ($\ll 1$); ν_0 is the mean frequency of the three resonance lines. It can readily be shown that the spin–lattice relaxation time T_1 of the system is $1/2W$.

The P terms are given by (Bloembergen, Purcell and Pound 1948)

$$P_{m\to m-1} = P_{m-1\to m} = \tfrac{1}{4}\gamma^2 H_1^2 g(\nu)(I+m)(I-m+1), \qquad \ldots\ldots(4)$$

where, as is customary, γ is the nuclear gyromagnetic ratio, $2H_1$ is the amplitude of the radio-frequency magnetic field and $g(\nu)$ is a normalized line-shape function. It is therefore convenient to write

$$P_{-\frac{1}{2}\to-\frac{3}{2}} = P_{-\frac{3}{2}\to-\frac{1}{2}} = 3p_{-1}W; \quad P_{-\frac{1}{2}\to\frac{1}{2}} = P_{\frac{1}{2}\to-\frac{1}{2}} = 4p_0W; \quad P_{\frac{3}{2}\to\frac{1}{2}} = P_{\frac{1}{2}\to\frac{3}{2}} = 3p_1W, \qquad \ldots\ldots(5)$$

where $p_i = \frac{1}{2}\gamma^2 T_1\{H_1^2 g(\nu)\}_i$. At the centre of each line $g(\nu) = 2T_2$ and $p = \gamma^2 H_1^2 T_1 T_2$.

When the nuclear spin system is in a steady state all $dN_m/dt = 0$. In this condition equations (1), with (2), (3) and (5) yield solutions

$$\left.\begin{array}{l} N_{\frac{3}{2}} - N_{\frac{1}{2}} = \dfrac{N_{\frac{1}{2}}\Delta}{1+p_1} \simeq \dfrac{n_0}{1+p_1} \qquad N_{\frac{1}{2}} - N_{-\frac{1}{2}} = \dfrac{N_{-\frac{1}{2}}\Delta}{1+p_0} \simeq \dfrac{n_0}{1+p_0} \\ \qquad N_{-\frac{1}{2}} - N_{-\frac{3}{2}} = \dfrac{N_{-\frac{3}{2}}\Delta}{1+p_{-1}} \simeq \dfrac{n_0}{1+p_{-1}} \end{array}\right\} \qquad \ldots\ldots(6)$$

where $n_0 = \frac{1}{4}N\Delta$, and $N = \Sigma N_m$ is the total number of nuclei. We see from equations (6) that if the relaxation mechanism is magnetic in origin the application of power at any one line frequency will affect only the population difference controlling the strength of that line, the other differences being unaffected. Thus the saturation factor is $(1+p_i)^{-1}$ for each line, irrespective of what power, if any, is applied to the other lines. If any two lines are superposed the saturation of the combined line is the same, apart from a trivial difference which arises if the line shapes are not quite identical.

If on the other hand the relaxation process is predominantly quadrupolar in origin we can neglect the magnetic transitions brought about by the lattice. Following Pound (1950) the relaxation transition probabilities can then be written

$$\left.\begin{array}{ll} W_{\frac{3}{2}\to\frac{1}{2}} = W_{-\frac{1}{2}\to-\frac{3}{2}} = W_{\frac{3}{2}\to-\frac{1}{2}} = W_{\frac{1}{2}\to-\frac{3}{2}} = W, & W_{\frac{1}{2}\to-\frac{1}{2}} = W_{-\frac{1}{2}\to\frac{1}{2}} = 0, \\ W_{\frac{1}{2}\to\frac{3}{2}} = W_{-\frac{3}{2}\to-\frac{1}{2}} = W(1+\Delta), & W_{-\frac{1}{2}\to\frac{3}{2}} = W_{-\frac{3}{2}\to\frac{1}{2}} = W(1+2\Delta). \end{array}\right\} \qquad \ldots\ldots(7)$$

In a steady state all $dN_m/dt = 0$ and equations (1) with (5) and (7) yield solutions

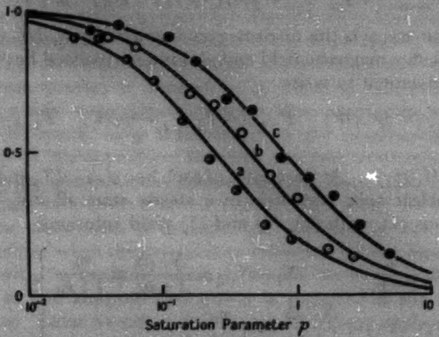
$$\left.\begin{aligned} &N_{-\frac{1}{2}} - N_{-\frac{3}{2}} = n_0(1+6p_0)(1+\tfrac{3}{2}p_1)A; \quad N_{\frac{1}{2}} - N_{-\frac{1}{2}} = n_0(1+\tfrac{15}{4}p_{-1}+\tfrac{15}{4}p_1+9p_{-1}p_1)A; \\ &\qquad\qquad N_{\frac{3}{2}} - N_{\frac{1}{2}} = n_0(1+6p_0)(1+\tfrac{3}{2}p_{-1})A; \\ &\text{where} \\ &\quad A \equiv [1+\tfrac{9}{4}p_{-1}+4p_0+\tfrac{9}{4}p_1+6p_{-1}p_0+6p_0p_1+\tfrac{9}{2}p_{-1}p_1+9p_{-1}p_0p_1]^{-1}. \end{aligned}\right\} \quad \ldots\ldots(8)$$

A number of cases will be considered.

(*a*) If power is applied only at the frequency of the transition between the upper sub-levels $(-\frac{1}{2}, -\frac{3}{2})$ then $p_0 = p_1 = 0$ and we get

$$\frac{N_{-\frac{1}{2}} - N_{-\frac{3}{2}}}{n_0} = \frac{1}{1+\frac{9}{4}p_{-1}}; \quad \frac{N_{\frac{1}{2}} - N_{-\frac{1}{2}}}{n_0} = \frac{1+\frac{15}{4}p_{-1}}{1+\frac{9}{4}p_{-1}}; \quad \frac{N_{\frac{3}{2}} - N_{\frac{1}{2}}}{n_0} = \frac{1+\frac{3}{2}p_{-1}}{1+\frac{9}{4}p_{-1}}. \quad \ldots\ldots(9)$$

We see therefore that the saturation factor for this satellite line is $(1+\frac{9}{4}p_{-1})^{-1}$, and this is plotted as curve *a* in the figure. We also see, as Pound showed, that when $p_{-1} \gg 1$ the centre line population difference is enhanced by a factor 5/3, and that of the other satellite line is depressed by a factor 2/3.



Saturation curves for nuclei with $I = 3/2$ in a crystal in which the operative spin–lattice relaxation mechanism is quadrupolar. Experimental points and theoretical relation are given for a satellite line by half-filled circles and curve *a*; for centre line by open circles and curve *b*; for centre line and one satellite superposed by filled circles and curve *c*. The ordinate gives relative values of the absorption part of the nuclear susceptibility.

If power is applied at the frequency of the other satellite line only $(p_{-1} = p_0 = 0)$ the behaviour is exactly the same, and is given by equations (9) with appropriate changes of suffixes.

(*b*) If power is applied only at the frequency of the centre line (transition $\frac{1}{2}$, $-\frac{1}{2}$), then $p_{-1} = p_1 = 0$ and we get

$$\frac{N_{-\frac{1}{2}} - N_{-\frac{3}{2}}}{n_0} = \frac{1+6p_0}{1+4p_0}; \quad \frac{N_{\frac{1}{2}} - N_{-\frac{1}{2}}}{n_0} = \frac{1}{1+4p_0}; \quad \frac{N_{\frac{3}{2}} - N_{\frac{1}{2}}}{n_0} = \frac{1+6p_0}{1+4p_0}. \quad \ldots\ldots(10)$$

We see therefore that the saturation factor for the centre line is $(1+4p_0)^{-1}$ and this is plotted as curve *b* in the figure. Curves *a* and *b* are identical when plotted logarithmically, apart from a lateral shift corresponding to a factor 9/16. In fact if the resonance lines have the same shape a given degree of saturation of the centre line should require only 9/16 of the power required to produce the same degree of saturation in the satellite. We also see that when $p_0 \gg 1$ the satellite population differences are both enhanced by a factor $\frac{3}{2}$.

(*c*) Now consider the situation when the centre line and the upper satellite are superposed and power is applied to this combined line only. (Second order perturbations usually prevent all three lines coalescing at any crystal orientation.) Then $p_{-1}=p_0$, $p_1=0$, and equations (8) become

$$\left.\begin{aligned} &\frac{N_{-\frac{1}{2}}-N_{-\frac{3}{2}}}{n_0}=(1+6p_0)B; \qquad \frac{N_{\frac{1}{2}}-N_{-\frac{1}{2}}}{n_0}=(1+\tfrac{15}{4}p_0)B;\\ &\qquad\qquad \frac{N_{\frac{3}{2}}-N_{\frac{1}{2}}}{n_0}=(1+6p_0)(1+\tfrac{3}{2}p_0)B;\\ &\text{where} \qquad\qquad B\equiv[1+\tfrac{25}{4}p_0+6p_0^2]^{-1}. \end{aligned}\right\} \qquad \ldots\ldots(11)$$

The power absorbed is proportional to $(N_{-\frac{1}{2}}-N_{-\frac{3}{2}})P_{-\frac{3}{2}\leftrightarrow-\frac{1}{2}}+(N_{\frac{1}{2}}-N_{-\frac{1}{2}})P_{\frac{1}{2}\leftrightarrow-\frac{1}{2}}$ and therefore proportional to $(1+\frac{33}{7}p_0)(1+\frac{25}{4}p_0+6p_0^2)^{-1}$. This saturation factor is plotted as curve *c* in the figure. It is closely similar to curves *a* and *b* but is displaced by factors of about 1·8 and 3·2 with respect to curves *a* and *b*. Since the saturation of the centre line alone enhances the population difference of either satellite and vice versa, it is not surprising that the combined line requires more power to produce a given degree of saturation than either line separately.

(*d*) Suppose the crystal is so oriented that the two satellite lines are superposed, and power is applied to this combined line only. Then $p_{-1}=p_1$, $p_0=0$, and equations (8) become

$$\frac{N_{-\frac{1}{2}}-N_{-\frac{3}{2}}}{n_0}=\frac{N_{\frac{3}{2}}-N_{\frac{1}{2}}}{n_0}=\frac{1}{1+3p_1}; \qquad \frac{N_{\frac{1}{2}}-N_{-\frac{1}{2}}}{n_0}=\frac{1+6p_1}{1+3p_1}. \qquad \ldots\ldots(12)$$

The saturation factor for the combined line is therefore $(1+3p_1)^{-1}$. This is not plotted in the figure, but is a curve geometrically similar to curves *a* and *b* lying exactly half-way between them.

Also shown in the figure are experimental saturation curves for the ^{23}Na resonance in a single crystal of sodium nitrate obtained at room temperature at 7·1 Mc/s for cases (*a*), (*b*) and (*c*); case (*d*) could not be studied because the combined satellites were not fully resolved from the centre line. The experimental points are in quite good agreement with the corresponding curves *a*, *b*, *c*, showing that for this crystal the relaxation mechanism is predominantly quadrupolar. As we saw earlier if the mechanism had been predominantly magnetic all the experimental points should have fallen on the same saturation curve $(1+p)^{-1}$. Pound (1950), using his two-coil method, also concluded that the relaxation mechanism was quadrupolar in a crystal of sodium nitrate.

The experimental results were obtained using a signal generator and twin-T bridge followed by amplification which included a phase-sensitive amplifier, from which the derivative of the spectrum was registered on a pen recorder. The values of p were obtained from relative values of the signal generator power to which H_1^2 is proportional, assuming $g(\nu)$ remains constant at the derivative turning points. Since the saturation conditions are those of case 2 (Bloembergen, Purcell and Pound 1948) the derivative maximum gives a relative measure of the absorbed power provided $g(\nu)$ remains constant. Actually however the positions of the turning points are slightly affected by saturation so that $g(\nu)$ does not remain quite constant. However, an analysis for lines with either Gaussian or Lorentz shape shows that neglect of this point is unimportant. A correction was however made to account for small differences in the widths of the lines.

It has been found that the relaxation time T_1 is slightly anisotropic. Therefore when the centre line alone was saturated the crystal was turned only a few degrees from the angular setting at which the centre line and a satellite were superposed.

REFERENCES

BLOEMBERGEN, N., PURCELL, E. M., and POUND, R. V., 1948, *Phys. Rev.*, **73**, 679.
POUND, R. V., 1950, *Phys. Rev.*, **79**, 685.

LETTERS TO THE EDITOR

Domain Patterns on Ferrite Single Crystals

Using well-known techniques (Bates 1951) powder patterns have been obtained on single crystals of some recently developed ferrimagnetic compounds, $BaFe_{12}O_{19}$ and $Ba_3Zn_2Fe_{24}O_{41}$, known as BaM and ZnZ respectively. BaM is otherwise known as ' magnadur ' and its properties have been fully described by Went *et al.* (1952), but the existence of ZnZ has only recently been reported by Jonker *et al.* (1956).

The crystal structures of these new compounds are in the hexagonal system and are related to the magnetoplumbite structure. The crystal anisotropy is rather high ($K \sim 3 \times 10^6$ erg cm^{-3}), and is such that the direction of spontaneous magnetization is parallel to the hexagonal axis.

Small crystals approximately 4 or 5 mm in cross section and thickness 0·1 or 0·2 mm were cleaved in the basal plane, so as to provide flat surfaces of high reflecting power. Drops of magnetic colloid were placed on these surfaces and with small applied fields normal to the surface, patterns of the type shown in figures 1, 2 and 3 (Plates I and II) were obtained. Figure 1 was taken with ZnZ and figures 2 and 3 with BaM. In figure 2 the thickness of the crystal decreased from about 0·2 mm in region A to approximately 0·1 mm in region B.

An interpretation of these patterns may be made following the general lines set out by Goodenough (1956) who considered the type of surface pole configurations to be expected in materials possessing high uniaxial anisotropy. The patterns are seen to vary with the thickness of the crystal, shown by figure 2 where the specimen thickness varied over the field of view and, whereas Sixtus *et al.* (1956) showed that a simple rick-rack pattern is formed on very thin crystals ($<10^{-3}$ mm), in thicker crystals a more complete pattern predominates with flower-shaped domains of reverse magnetization extending from the surface inwards.

It should prove possible to obtain patterns on progressively thinner plates of the crystal and if necessary observe the depth of the reverse domains by looking at the patterns on the side surfaces of the specimen.

The crystals used in the present work were kindly supplied by the Philips Research Laboratories, Eindhoven.

Mullard Research Laboratories,
Cross Oak Lane,
Salfords, nr. Redhill, Surrey.
24th January 1957.

R. F. Pearson.

Bates, L. F., 1951, *Modern Magnetism* (Cambridge : University Press).
Goodenough, J. B., 1956, *Phys. Rev.*, **102**, 356.
Jonker, G. H., Wijn, H. P. J., Braun, P. B., 1956, *I.E.E. Convention on Ferrites.*
Sixtus, K. J., Kroenberg, K. J., Tenzer, R. K., 1956, *J. Appl. Phys.*, **27**, 1051.
Went, J. J., Rathenau, G. W., Gorter, E. W., G. W. van Ousterhout, 1952, *Philips Tech. Rev.*, **13**, No. 7, 194.

REVIEWS OF BOOKS

Roentgenstrukturanalyse von Kristallen, by R. KOHLHAAS and H. OTTO. Pp. 212. (Berlin : Akademie–Verlag, 1955.) DM 23.

" To learn and at due times to repeat what one has learnt, is that not after all a pleasure ? " is the opening sentence of *The Analects*. When this represented the object of study only Four Books and Five Classics made up the corpus of official learning but today, even in a restricted field, it is manifestly impossible for a student to work through the whole literature of a subject. Besides being voluminous, the material is continually changing, parts are becoming obsolete and parts are expanding in new directions. The preparation of an elementary textbook is therefore an increasingly responsible and difficult task as it is the duty of the teacher to expound, not only the anatomy and morphology of the corpus, but its embryology, physiology and evolution. To do this in a limited space requires an exceptional concentration on principles and outlines to show the structure and operation of the subject.

The reviewer has hopes of seeing an inspiring and suggestive textbook on crystallography to serve as an introduction to the more specialized and advanced texts such as Volume 3 of *The Crystalline State* and to replace Volume 1 of *The Crystalline State* which was written more than 23 years ago.

However, Kohlhaas and Otto's book is on traditional lines and is a careful and competent presentation of certain well-established parts of crystallography. It achieves brevity chiefly by omission, referring the reader to other books for the supplementary topics.

The chief criticisms are of lack of awareness of current literature and practice. For example, the Schiebold–Sauter camera is given much space when few crystallographers have even seen one, while Buerger's camera for giving symmetry-true photographs (it did not in fact do that exactly) is confused with the precession camera developed from it, and the latter, although widely used, is not discussed. Mention is made of only the original de Jong and Bouman camera and not of subsequent improved versions. The illustrative structure analyses are well chosen and described, although the notation might be modernized. The general presentation, particularly of some of the three tone diagrams, is good, but misprints occur. One Laue photograph (Figure 117, below) shows a poor technique which should not be perpetuated.

On the whole, the book seems more suitable as an introductory text for crystallographers who intend to practise, rather than as an account for chemists and geologists of what crystallography is about. Attention is chiefly concentrated on the geometry of crystal lattices and of diffraction cameras. It would be unfortunate if chemists were left with the impression that when the intensities have been measured and corrected a structure determination is nearly finished. The material is too detailed for non-specialists and too limited in extent for specialists. In the German literature there is no doubt a place for this book, but the field is already well covered in English.

A. L. MACKAY.

Advances in Electronics and Electron Physics, Volume VII. Edited by L. MARTON. Pp. x + 525. (New York: Academic Press, 1956.) $11.50.

The great flood of publications, good, bad and indifferent, in any single branch of science is almost beyond the capacity of any one individual to read, let alone to digest and assimilate if he wishes to do any research and teaching himself at the same time. The solution to this dilemma seems to be the regular publication of critical reviews by specialists of advances in their own fields at the level that can be readily followed by non-specialists and with comprehensive bibliographies to assist in more specialized reading.

The series of volumes published by the Academic Press, the titles of which all begin with the words: *Advances in* admirably fulfil this requirement and they cover several fields, other than that with which we are concerned here, such as *Applied Mechanics* and *Carbohydrate Chemistry*.

The series, of which the volume under review is the seventh, began as *Advances in Electronics* but as from Volume VI the title has been widened to: *Advances in Electronics and Electron Physics* to cover more adequately the scope of the review articles now being included.

The field, once fairly adequately covered by the title *Electronics*, has expanded so rapidly during the last couple of decades that the Editor of these review volumes must have found it difficult to choose from such a wealth of possible subjects. It is to his great credit that both subjects and authors have been chosen so skilfully. Almost half of this present volume is devoted to those aspects of solid state physics which have become so important in electronics—almost entirely of post-war development. Two more chapters, *Observational Radio Astronomy* and *Analogue Computers* are on subjects which have practically no pre-war history. Only two chapters: *Sputtering by Ion Bombardment* and *Electrical Discharges in Gases and Modern Electronics* have any appreciable pre-war bibliography.

The present reviewer admits his inability seriously to criticize the subject matter of these chapters. Perhaps it is sufficient to say that he finds them excellent for acquiring an up-to-date working knowledge of each field such as might be expected of a post-graduate student about to begin research in such a field. Then, for a more comprehensive knowledge, the excellent bibliographies will be of great assistance.

This series is undoubtedly doing a most important task and Editor and publishers are to be congratulated on their success in persuading so many outstanding specialists to summarize progress in their own fields for the general benefit. J. D. MCGEE.

Radiology Physics, by J. K. ROBERTSON. Pp. xvi + 330. (London: Macmillan, 1956.) 36*s*.

This is the revised third edition of a book which was originally published in 1940 and again in 1945 in revised form. The book is intended to provide medical students, radiologists and radiographers with an elementary account of the physical principles involved in radiology and the applications of radioactive isotopes in medicine. This is a very difficult job for any writer to attempt but the author, who was formerly Professor of Physics at Queen's University, Kingston, Canada, has great experience in teaching medical students and is therefore well qualified to undertake such a task.

The first six chapters, of the sixteen chapters making up the book, give the student a general introduction to alternating current, the production of high voltages, cathode rays, x-ray tubes, rectification and the use of valves. These chapters cover their subjects adequately, assuming a little previous knowledge of electricity and some elementary mathematics. The only comment to be made about these chapters concerns the title of Chapter V. At the beginning of the chapter the title is *Roentgen Tubes* but the title given at the head of each page in the chapter is *Roentgen Rays*. From the subject matter of the chapter it appears that the former is the correct title.

The serious criticism of the book applies to the chapters from seven onwards in which the general properties of x-rays and their interaction with matter are discussed. These basic principles must be expounded to the student clearly and in logical sequence if he is to understand the techniques of radiology. The book falls short in this important respect. For example in the chapter dealing with secondary x-rays and absorption, the author obviously thinks that a straightforward and simple explanation in terms of the three main processes of photoelectric effect, Compton effect and pair production is too difficult. Instead, the reader goes through sections in this chapter with the titles of *Secondary X-rays*, *Scattered Radiation*, *Characteristic Radiation*, *Recoil Electrons* and *Scattering*. In this chapter pair production is dismissed in two lines and the reader is referred to Chapter 15 for further information. On referring to Chapter 15, which is entitled *Artificial Radioactivity*, a short account of pair production is found and by way of illustration a cloud chamber photograph is shown. For those who are familiar with cloud chamber photographs it is indeed a good illustration but for medical students or radiologists it is quite useless since little or no explanation of the photograph is given.

When this book was first published in 1940 it was common practice in radiology for an x-ray photon to be denoted by its wavelength. However, in recent years the x-ray photon has usually been referred to by its energy. This change in terminology is necessary for the sake of uniformity throughout the very large range of x-ray energies now used and also to keep in line with nuclear physics where the γ-ray is always denoted by its energy. The use of photon energy can also be justified on the grounds that the interaction of x-rays with matter can be explained more simply in terms of energy change rather than wavelength change. In *Radiology Physics* both terminologies are used, x-ray wavelengths in the older sections and x-ray energy in most of the new material of the third edition. This mixture of terminologies should have been avoided by the author even though it would have meant more drastic revision of the older material.

The new material added to chapter 13 which deals with the production of very high voltages contains some doubtful statements. For example it is stated that 400 Mev electrons from a betatron have been used at the University of Illinois for the treatment of superficial carcinoma of the shin. This seems an unlikely procedure but it is possible that 400 Mev protons were used, in which case they would not be produced by a betatron. Mayneord is attributed with the statement that the conversion efficiency of 20 Mev electrons into x-rays in a betatron is as high as 70%. This is the expected efficiency for a very thin target but the reader should be told that the efficiency of a thick target is only about 20% for 20 Mev electrons. Finally, the author adds a little hearsay at the end of the chapter by remarking that a 600 Mev proton linear accelerator has been used at Harwell. A machine of this type was contemplated about two years ago but as far as it can be ascertained the project has since been dropped.

The last three chapters of the book deal with the elementary nuclear physics necessary for understanding the use of radioactive isotopes in medicine. Medical students will find these chapters particularly useful in supplementing their lectures on radioactivity.

The general impression is that this was a good textbook in its first edition in 1940 but the revised third edition in 1956 leaves much to be desired. Rapid advances have taken place in the field of physics applied to radiology since 1940 and a text book on the subject cannot be kept up to date by the addition of new material to its original text. At some time an entirely new book is required and it is thought that this would have been an appropriate time. Apart from this rather severe criticism the other general point of criticism is that the text could have been more fully illustrated by diagrams and a few well-tried photographs together with some cloud chamber photographs might have been omitted. The good point about the book is the selection of topics and in the brevity with which they are treated thus avoiding overloading a student on his first reading.

G. W. DOLPHIN.

The Interference Systems of Crossed Diffraction Gratings, by J. GUILD. Pp. viii + 152. (Oxford: Clarendon Press, 1956.) 25*s*.

This book describes the conditions under which moiré fringes may be formed with the aid of two crossed grating structures. Although patterns of this general type are frequently to be seen around us it is only within recent years that new methods of manufacture have made available gratings of sufficient size and quality to make the fringes of use in engineering practice. It is to be noted that Mr. Guild has for long been closely associated with the manufacture and testing of these large gratings, and also that this work represents the first detailed account of the formation of moiré fringes.

The primary purpose of the book is to supply the theoretical background required by the technologist who is contemplating the use of moiré fringes for linear measurement, and there is no attempt to discuss the practical details of particular systems. The first two chapters form an introduction to the main work and include a derivation of the fundamental grating formula. There is to be found here a carefully prepared section on terminology which one would do well to record in sketch form for future reference. The formation of the fringes produced by the mutual interaction of the spectra given by the two gratings in series is described in the next chapter and results in a cosine series for the distribution of intensity in the moiré pattern. The effect of the amplitude and phase of the terms of this series on the pattern is then discussed with particular reference to the two extreme cases of multiple beam and two-beam interference. The importance of using a spectroscopic method of observation is stressed and the effect of finite slit width and finite grating separation on the fringe contrast is included, together with an account of the coloured patterns that can be produced if a white source is employed. A simple practical demonstration of the effect of the finite aperture of the eye on resolution should prove helpful to those who are new to the subject. There follows a regrettably short chapter on the actual problems of using moiré fringes in metrology but we are assured that a separate paper dealing more fully with the practical aspects is in course of preparation.

The content of the last chapter, which deals with the influence of characteristic ruling errors on the linearity, slope and periodicity of the fringes is well illustrated by a number of excellent plates and should be of particular interest to those concerned with grating manufacture and testing.

The essentially geometrical treatment used throughout may disappoint some students of optics who would prefer to regard the subject as a problem in the diffraction theory of image formation. It is not easy to see from the treatment given here, for example, how the distribution of intensity in the fringes is directly influenced by the ruling profile, by the coherence of the incident wave over both gratings, and by the aberrations and finite aperture of the imaging lens. Bearing in mind, however, that the book is primarily intended for the technologist it must be admitted that most of the optical problems involved in the use of the gratings now available have been covered. Although the mathematical notation is a little tedious in places the text itself is fairly easy to follow, and since the book assumes little previous knowledge of optics and contains the results of interest in practice it should be welcomed by those contemplating the use of moiré fringes for precise linear measurement and machine tool control. L. R. BAKER.

Glass, by G. O. JONES. Pp. vi + 119. (London: Methuen's Monographs on Physical Subjects; New York: John Wiley, 1956.) 8*s*. 6*d*.

Professor Jones' book affords a most valuable introduction to the study of the glassy state, more particularly for those trained in the pure sciences than for the intending glass technologist. At the outset a definition of glass is adopted which extends the discussion beyond the bounds of conventional technical materials and focuses the reader's attention on the problem of establishing an atomic model from which accurate deductions of the distinctive properties of a glass could be made. However, in the absence of a quantitative model for even simple liquids our ideas about associated liquids and glasses are rudimentary and qualitative, in marked contrast to the theory of crystalline solids. The situation is revealed clearly by this compact summary which presents the study of glass as a challenge to the theoretician.

Throughout the book atomic structure is discussed thoroughly but the variation of properties with composition receives scant attention, to the disappointment, no doubt, of glass technologists. An excellent discussion of the transformation range phenomena is given, including a formal thermodynamic treatment based on the idea of fictive temperature and the validity of this idea is also discussed. The rheological properties are described clearly and the complex data on strength are ably summarized. Only small sections are devoted to optical and electrical properties on the grounds that these differ little in essentials between glasses and crystalline solids. The chapter devoted to crystallization and the conditions for glass formation rather oversimplifies the relation between viscosity and devitrification; however, the difficulty of establishing reliable rules for glass-forming compositions is amply demonstrated and some of the compositions quoted throughout the book show how wide is the field of search for new glasses.

In general, a great deal has been condensed into a small book with clarity, but it is a pity that so many footnotes have been interspersed as most of them could have been incorporated in the text without confusion. J. O. ISARD.

BOOK NOTICES

Sunspot and Geomagnetic-Storm Data derived from Greenwich Observations, 1874–1954. Compiled under the direction of Sir Harold Spencer Jones. (London: H.M. Stationery Office, 1955.) 25*s.*

The Relations between Scientific Research in the Universities and Industrial Research (A Report on Conditions in Great Britain), Edited by V. E. Cosslett. Pp. vii + 187. (London: International Association of University Professors and Teachers, 1955.) 15*s.*

Paris: Publications Scientifiques et Techniques du Ministère de l'Air.

Les applications scientifiques de la notion de régularité, par P. Vernotte. Pp. 84. No. 307, 1956. 750 *fr.*

Étude expérimentale des écoulements gazeux, par M. Merle. Pp. 89. No. 308, 1956. 1,100 *fr.*

Nouvelle méthode approchée de calcul des couches limites laminaire et turbulente en écoulement compressible, par A. Walz. Pp. 87. No. 309, 1956. 950 *fr.*

Étude de la lubrification sous pression rapidement variable. Cas des lubrifiants plastiques. Existence de leur anisotropie. Par F. Charron. Pp. 27. No. 310, 1956. 450 *fr.*

Étude éxperimentale de la convection forcée sur paroi mobile (*Hyperconvection favre*), par C. Bory. Pp. 27. No. 312, 1956. 1000 *fr.*

Utilisation des compteurs de Geiger-Müller à l'étude de variations de texture de laminage suivant le température de déformation, par F. Provost. Pp. 57. No. 313, 1956. 800 *fr.*

Sur la turbulence spatialement homogène d'un fluide compressible, par M. Z. Krzywoblocki. Pp. 56, No. 314, 1956. 650 *fr.*

Contribution a l'étude des phénomènes de durcissement dans l'alliage aluminium–cuivre a 4% de cuivre, par R. Graf. Pp. 100. No. 315, 1956. 1,100 *fr.*

Notes Techniques:

Étude de la diffusion dans les gels à l'aide des isotopes radioactifs, par S. Cordier. Pp. 72. N.T. 57. 450 *fr.*

Sur un curieux cas d'alternance de tourbillons, par E. A. Brun. Pp. 27. N.T. 58, 1956. 350 *fr.*

CORRIGENDUM

The Statistics of Divalent Impurities in a Semiconductor, by C. H. Champness (*Proc. Phys. Soc. B*, 1956, **69**, 1335).

(1) Delete the factor 2 in the third term of the cubic equation on p. 1337.

(2) Delete "and when the two levels are well separated" in the last sentence on p. 1337 and insert after the first sentence on p. 1338—"Suppose the two levels are well separated ($\alpha \gg \beta$)".

CONTENTS OF SECTION A

PAGE

Dr. F. A. el Bedewi and Mr. I. Hussein. Magnetic Analysis of Tritons and Alpha Particles Emitted from the Deuteron Bombardment of Teflon . . 233

Dr. I. C. Percival. Electron Excitation of 3p Levels of O I 241

Dr. J. C. Barton. The First Passage Problem in the Theory of Multiple Scattering 253

Prof. J. A. V. Fairbrother, Mr. D. G. Parkyn and Mr. B. M. O'Connor. The K Auger Yield for Tin 262

Dr. A. D. Brailsford. The Magnetic Energy Levels of Electrons in Metals . 275

Dr. R. Marriott. The Conversion of Metastable Helium from the Singlet to the Triplet State by Electron Collisions 288

Dr. R. Marriott and Dr. M. J. Seaton. A Simple Analytic Wave Function for He 1s2s ^{1}s 296

Dr. W. F. Miller and Prof. R. L. Platzman. On the Theory of the Inelastic Scattering of Electrons by Helium Atoms 299

Research Notes:

Mr. S. A. Husain and Mr. J. L. Putman. The Absorption of Beta Particles as a Function of Atomic Number 304

Prof. D. R. Bates and Mr. A. Williams. Inelastic Collisions between Heavy Particles. VII: Electron Loss from Fast Hydrogen Atoms passing through Helium 306

Prof. W. E. Burcham. Light Output of Potassium Iodide Crystals under Bombardment by Carbon Ions 309

Letters to the Editor:

Mr. L. S. Kothari and Dr. K. S. Singwi. Thermal Inelastic Scattering of Cold Neutrons in Polycrystalline Graphite 314

Reviews of Books 316

Contents of Section B 320

Figure 1. Patterns on the basal plane for ZnZ.

Figure 2. Patterns on the basal plane for BaM.

Figure 3. Patterns on the basal plane for BaM.

The Aberration Permissible in Optical Systems

By H. H. HOPKINS
Department of Physics, Imperial College, London

MS. received 12th September 1956, and in final form 29th January 1957

Abstract. The tolerance on aberration is made subject to the criterion that the response of the optical system for a given spatial frequency shall not be less than 0·8 of the response of the same system in the absence of aberration. In this way the tolerance on aberration is related to the number of lines to be resolved with good contrast. The method is applied to defect of focus, spherical aberration, coma and astigmatism. The best form of correction and focal plane according to this criterion are determined by the same method. The results are expressed by approximate formulae for the lower spatial frequencies which are those of importance in other than very highly corrected systems.

---

§ 1. INTRODUCTION

THE analytical study of the influence of large aberrations on the diffraction images of point sources formed by optical systems has met with very limited success. Apart from results obtained by numerical integrations, and by the use of mechanical integrators (Maréchal 1948), only the influence of small wave-front aberrations, less than about 2λ, has been successfully investigated. A similar situation exists with regard to the diffraction theory of aberration tolerances. The criterion of Rayleigh, for example, stipulates that the asphericity of the image-forming wave front shall be such that the wave front lies wholly within two concentric spheres, centred on the chosen focus and differing in radius by $\lambda/4$. The Strehl intensity ratio (Definitionshelligkeit) for a given focal plane is defined to be the ratio at the greatest maximum of intensity in the diffraction image in that plane to the intensity obtained at the geometrical focus in the absence of aberration. The tolerance criterion implied by imposing the condition that the Strehl intensity ratio should not fall below 0·8 amounts in practice to almost the same condition as the $\lambda/4$ limit of Rayleigh. A comprehensive treatment of aberration tolerances based on the Strehl criterion has been given by Maréchal (1948).

An optical system satisfying the Rayleigh criterion gives an image quality little inferior to that of a perfect system, no matter what size of detail in the object is of interest. The Maréchal treatment of tolerances is therefore suitable for very highly corrected systems, such as microscope objectives. It is, however, well known in practice that optical systems having much larger aberrations are capable of producing images of good quality of objects in which the size of detail is coarse in comparison with the inherent limit of resolution. Recent theoretical studies (Hopkins 1955, De 1955) have shown to what extent this is true in the two cases of defective focusing and astigmatism. These studies have yielded formulae for tolerances on defective focusing and astigmatism which are functions of the size of detail which it is desired to reproduce with good contrast. These tolerance formulae are only valid for large aberrations, when only the lower spatial frequencies come in question. They were derived from the leading terms in analytical expressions for the so-called response functions of optical systems for the two cases. In what follows a general method is given for obtaining the permissible tolerance on aberration, based on the criterion that the response of the system for a given spatial frequency shall not fall below 0·8 of that of a perfect system for the same frequency. The method, which applies equally to large aberrations and the whole frequency range, is applied to the cases of defect of focus, astigmatism, spherical aberration and coma. In the first two cases the resulting formulae give forms identical with the earlier results of Hopkins (1955) and De (1955) for coarser detail. It will be convenient to summarize first the methods used for treating the problem of the frequency response of optical systems.

§ 2. General Formulae

In figure 1, Σ is a reference sphere of centre O′, and P′ is a point in a focal plane passing through O′ and perpendicular to the principal ray E′O′. α' is the angular aperture of a given reference point, such as B, on the reference sphere. In most cases it is convenient to take this point to be on the periphery of the image-forming pencil, so that α' measures the angular semi-aperture of the latter. If n' is the refractive index of the image space, the reduced rectangular coordinates of the point P′ are defined by the relations

$$u' = \frac{n' \sin \alpha'}{\lambda} \xi', \qquad v' = \frac{n' \sin \alpha'}{\lambda} \eta' \qquad \text{......(1)}$$

in which (ξ', η') are the geometrical coordinates of P′, and λ is the vacuum wavelength of light. Reduced coordinates (u, v) are defined for the object plane by expressions identical with (1), the different quantities being then those corresponding to the object space. The variables (u, v), (u', v') used here differ from those used in previous papers (Hopkins 1953, 1955, 1956, De 1955) by the omission of a factor 2π.

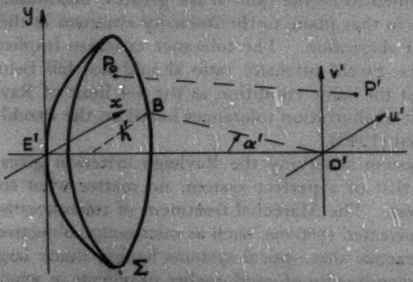


Figure 1. Reference sphere and image plane.

In a line structure whose elements are parallel to the v axis, the intensity is a function $B(u)$ of u alone. The image of such a line structure will have an intensity distribution $B'(u')$, which is likewise independent of v'. Provided the distributions of light in the images of parallel line sources formed in the neighbourhood of O_1' (figure 1) do not differ appreciably over distances which are at least as large as the region within which a single such image has appreciable intensity, the normalized intensity distribution in the image may be written

$$B'(u') = \int_{-\infty}^{+\infty} B(u)G(u'-u)\,du \qquad \ldots\ldots(2)$$

where $G(u')$ is the distribution of intensity in the image of a line source whose geometrical image is formed at $u'=0$. This is the condition known as isoplanatism. By virtue of the convolution integral (2), the inverse Fourier transforms of the three functions satisfy the product relation

$$b'(s) = g(s)b(s), \qquad \ldots\ldots(3)$$

the inverse Fourier transforms being defined by

$$b(s) = \int_{-\infty}^{+\infty} B(u)\exp\{-i2\pi us\}\,du, \qquad \ldots\ldots(4)$$

with similar expressions for $g(s)$, and $b'(s)$. The normalized frequency response is defined to be

$$D(s) = g(s)/g(0) \qquad \ldots\ldots(5)$$

in which the factor $1/g(0)$ merely implies a convenient choice of photometric units for the object and image planes respectively. This ratio is such that $D(0)=1$, and the object $B(u)=1$ gives rise to an image $B'(u)=1$.

If R is the number of cycles per unit length of a cosinusoidal component of the intensity distribution in the object plane, and R' the corresponding quantity for the image plane, the spatial frequency variable s is given by

$$s = \frac{\lambda}{n\sin\alpha}R = \frac{\lambda}{n'\sin\alpha'}R' \qquad \ldots\ldots(6)$$

since, in terms of the coordinates (1), the length of period is

$$u_0 = \left(\frac{n\sin\alpha}{\lambda}\right)\frac{1}{R},$$

and $u_0 s = 1$.

It is often convenient to write $D(s)$ in terms of a modulus and phase, namely

$$D(s) = T(s)\exp\{i\theta(s)\}. \qquad \ldots\ldots(7)$$

The distributions of intensity in the object and image planes are necessarily real. Writing for the object intensity

$$B(u) = 2\int_{0}^{+\infty} \beta(s)\cos\{2\pi us + \phi(s)\}\,ds, \qquad \ldots\ldots(8)$$

we note that inversion of (3) gives

$$B'(u') = 2\int_{0}^{\infty} T(s)\beta(s)\cos\{2\pi u's + \phi(s) + \theta(s)\}\,ds. \qquad \ldots\ldots(9)$$

The modulus $T(s)$ thus represents the ratio of the contrast in the image of a cosinusoidal intensity distribution, of spatial frequency s, to that in the object itself. Corresponding to the phase shift $\theta(s)$ is a lateral displacement of this component of the image distribution equal to $\delta u' = \theta(s)/2\pi s$.

To calculate the inverse Fourier transform $g(s)$ we utilize the fact that the distribution of complex amplitude in the image of a point source is given by the

Fourier transform of the pupil function. This latter is defined to be

$$\left.\begin{aligned} f(x,y) &= \exp\{ikW(x,y)\} \quad \text{within A} \\ &= 0 \quad \text{outside A,} \end{aligned}\right\} \qquad \ldots\ldots(10)$$

A being used to denote the region of the pupil, and $k = 2\pi/\lambda$. The function $W(x,y)$ expresses the asphericity of the wave front (wave aberration) as a function of position over the reference sphere Σ (figure 1). The axis of y is chosen to be parallel to the direction of constant intensity in the line object, and reduced coordinates

$$x = \frac{X}{h'}, \qquad y = \frac{Y}{h'} \qquad \ldots\ldots(11)$$

are employed, where (X, Y) are the geometrical coordinates of the current point P_0 on the reference sphere, and h' is the distance from the axis of the fixed point B, whose angular aperture α' is used in the definitions of (u', v') in (1). In this definition we suppose the image-forming wave fronts to be of uniform amplitude. The treatment of aberrational tolerances for non-uniform waves presents no additional difficulties, but the analysis is simplified if uniform amplitude is assumed, and this assumption is, moreover, valid in most cases of practical interest.

The aberrations of the optical system will usually be described with reference to rectangular coordinates (x_0, y_0), in which the y_0 axis lies in the meridian plane. If $W_0(x_0, y_0)$ is the wave front aberration with reference to (x_0, y_0), and the direction of the line-structure under consideration makes an angle ψ with the y_0 axis, the axes (x, y) are turned through an angle ψ relative to those of (x_0, y_0) and the aberration function to be used in (10) is

$$W(x,y) = W_0(x\cos\psi - y\sin\psi, \quad y\cos\psi + x\sin\psi). \qquad \ldots\ldots(12)$$

In practice, when $W_0(x_0, y_0)$ is written as a function of $x_0^2 + y_0^2$ and y_0, we merely require to replace these terms by $x^2 + y^2$ and $y\cos\psi + x\sin\psi$ respectively.

The frequency response of the optical system will generally vary with orientation of the line structure; to denote this fact we shall write $D(s, \psi)$ for the normalized response defined by (5).

With the notation described above, the complex amplitude in the diffraction image of a point source, situated at the point (0, 0) of the object plane, is described, apart from a constant factor, by the distribution

$$F(u', v') = \iint_{-\infty}^{+\infty} f(x,y) \exp\{i2\pi(u'x + v'y)\}\, dx\, dy \qquad \ldots\ldots(13)$$

and the distribution of intensity in the pattern is given by

$$G(u', v') = |F(u', v')|^2. \qquad \ldots\ldots(14)$$

The distribution of intensity in the image of a line source situated at $u = 0$ is

$$G(u') = \int_{-\infty}^{+\infty} G(u', -v)\, dv$$

and the inverse Fourier transform of this distribution is given by

$$g(s) = \iint_{-\infty}^{+\infty} G(u', v') \exp\{-i2\pi u's\}\, du'\, dv' \qquad \ldots\ldots(15)$$

in which, for convenience a prime has been added to the variable of integration v.

Using (13) and (14) the formula (15) is easily transformed to

$$g(s)=\iint_{-\infty}^{+\infty} f(x,y)f^*(x-s,y)\,dx\,dy,$$

the asterisk denoting a complex conjugate. The normalized frequency response (5) is thus

$$D(s,\psi)=\frac{1}{A}\iint_{-\infty}^{+\infty} f(x+\tfrac{1}{2}s,y)f^*(x-\tfrac{1}{2}s,y)\,dx\,dy \qquad \ldots\ldots(16)$$

where $A=g(0)$ is simply equal to the area of the pupil, and the origin has been shifted to the point $(\frac{1}{2}s, 0)$ in the (x, y) plane.

The pupil function $f(x, y)$ in (16) is that defined by (10). The region of the pupil, A, will usually be a circle for the axial part of the image. For extra-axial points the region A will then be an oblique conical projection of the circular exit pupil, and possibly further limited by vignetting. The factor $f(x+\frac{1}{2}s, 0)$ in (16) is zero for points outside a pupil region centred on the point $x=-\frac{1}{2}s, y=0$. Similarly, the factor $f^*(x-\frac{1}{2}s, 0)$ is zero for points outside a pupil centred on the point $x=+\frac{1}{2}s, y=0$. The effective region of integration in (16) is therefore that common to these two pupil areas. This region, shown in figure 2, we shall denote by S. For a line structure inclined at an angle ψ to the meridian plane, the two pupils whose overlap defines the region S have their centres at the points $x_0=\pm\frac{1}{2}s\cos\psi$, $y_0=\pm s\sin\psi$. It will be apparent that, in general, the region of integration S will depend on ψ, the orientation of the line structure considered. When the pupil is of circular form, the form of the region S, when referred to the (x, y) axes, is independent of ψ. It is convenient in this case to take the angle α' to be the angular semi-aperture of the image-forming cone of light, so that

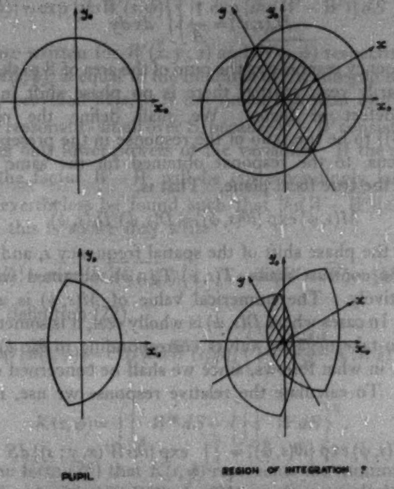


Figure 2. Non-circular and circular pupils.

$x^2+y^2=1$ is the periphery of the pupil. The two circles defining S are then of unit radius, and their centres are separated by a distance s. In what follows, the pupil will be assumed to be circular for both axial and extra-axial points of the image plane.

Substituting the pupil function (10) in (16), and using S to denote the region of integration, we find

$$D(s,\psi)=\frac{1}{A}\iint_S \exp\{iksW(x,y;s)\}\,dx\,dy \qquad \text{......(17)}$$

where

$$W(x,y;s)=\frac{1}{s}\{W(x+\tfrac{1}{2}s,y)-W(x-\tfrac{1}{2}s,y)\}$$

or, expanding each of the terms in the right-hand side of this latter about the point (x,y),

$$W(x,y;s)=W_x'(x,y)+\frac{1}{3!}\left(\frac{s}{2}\right)^2 W_x'''(x,y)+\frac{1}{5!}\left(\frac{s}{2}\right)^4 W_x^{\mathrm{V}}(x,y)+\ldots, \qquad \text{......(18)}$$

$W_x'(x,y)$, ..., being the successive derivatives of $W(x,y)$ with respect to x.

In the absence of aberration $W(x,y)=0$, and

$$D_0(s,\psi)=\frac{1}{A}\iint_S dx\,dy$$

expresses the frequency response as the ratio of the area of S to that of the pupil A. $D_0(s,\psi)$ is necessarily real, so that there is no phase shift in the absence of aberration and defect of focus. We shall define the relative response $M(s,\psi)\exp\{i\theta(s,\psi)\}$ to be the ratio of the response in the presence of aberration, and defect of focus, to the response obtained for the same system without aberration and in the true focal plane. That is,

$$M(s,\psi)\exp\{i\theta(s,\psi)\}=D(s,\psi)/D_0(s,\psi), \qquad \text{......(19)}$$

in which $\theta(s,\psi)$ is the phase shift of the spatial frequency s, and $M(s,\psi)$ is equal to the ratio of the contrast ratios $T(s,\psi)/T_0(s,\psi)$, obtained with and without aberration respectively. The numerical value of $M(s,\psi)$ is always less than or equal to unity. In cases where $D(s,\psi)$ is wholly real, it is sometimes convenient to allow $M(s,\psi)$ to take negative values corresponding to $\theta(s,\psi)=(2n+1)\pi$, but this will not occur in what follows, since we shall be concerned only with values of $M(s,\psi)\geqslant 0{\cdot}80$. To calculate the relative response we use, in place of (17), the form

$$M(s,\psi)\exp\{i\theta(s,\psi)\}=\iint_S \exp\{iksW(x,y;s)\}\,dS \qquad \text{......(20)}$$

dS being a reduced element of area $dS=dx\,dy/S$, S being the area of the region of integration.

In the above a line structure has been assumed as a typical test-object, and the object and image planes have been assumed to be perpendicular to the principal ray in the object and image space respectively. It is easily shown (Hopkins 1956) that the response for a frequency pair (s_0,t_0) of a two-dimensional object is identical with that for the frequency $s=\sqrt{(s_0^2+t_0^2)}$ of a line structure inclined at an angle $\psi=\tan^{-1}(t_0/s_0)$ to the t_0 axis. In fact any two-dimensional object may be regarded as the superposition of a set of line structures having different orientations ψ. If the image is observed in an image plane perpendicular to the optical axis, the response for the spatial frequency (s_0, ψ_0) may be shown (Hopkins 1956) to be equal to that for the frequency

$$s = s_0\{\cos^2\psi_0 + \sec^2\beta' \sin^2\psi_0\}^{1/2}, \quad \psi = \tan^{-1}(\sec\beta' \tan\psi_0) \quad \ldots\ldots(21)$$

observed in the image plane perpendicular to the principal ray. With these relations at hand the restrictions imposed in the present work do not limit the application of the results.

§ 3. TOLERANCE CRITERION AND FORMULAE

In (20) the numerical range of variation of the argument will be reduced, and the approximations used later improved, if a factor $\exp\{iks\overline{W(s,\psi)}\}$ is taken outside the integral sign, where

$$\overline{W(s,\psi)} = \iint_S W(x,y\,;s)\,dS \quad \ldots\ldots(22)$$

is the mean value of $W(x,y\,;s)$ over the region S. The expression (20) then takes the form

$$M(s,\psi)\exp\{i\theta(s,\psi)\} = \exp\{iks\overline{W(s,\psi)}\}\iint_S \exp\{iks(W-\overline{W})\}\,dS \quad \ldots\ldots(23)$$

in which W, $\overline{W}$ are written for $W(x,y\,;s)$ and $\overline{W(s,\psi)}$ respectively. The value of the modulus, $M(s,\psi)$, is equal to the modulus of the integral on the right-hand side of (23); and, provided $M(s,\psi)$ is not much less than unity, the argument $ks(W-\overline{W})$ will be reasonably small over S, because $W-\overline{W}$ consists of a polynomial in (x,y) containing only lower powers of the variables. If the aberration of the system is large, the factor $W-\overline{W}$ will be correspondingly large, but a small value of s may nevertheless be found such that $|ks(W-\overline{W})| \leqslant 1$ for the whole region S. When this is so we may write

$$M(s,\psi)\exp\{i\theta(s,\psi)\} = \exp\{iks\overline{W(s,\psi)}\}\iint_S \{1 + iks(W-\overline{W}) - \tfrac{1}{2}k^2s^2(W-\overline{W})^2\}\,dS$$

or, in view of the definition (22),

$$M(s,\psi)\exp\{i\theta(s,\psi)\} = \exp\{iks\overline{W(s,\psi)}\}\left\{1 - \frac{2\pi^2s^2}{\lambda^2}K(s,\psi)\right\} \quad \ldots\ldots(24)$$

where

$$K(s,\psi) = \iint_S W^2\,dS - \left\{\iint_S W\,dS\right\}^2. \quad \ldots\ldots(25)$$

It is clear from the form (25) that $K(s,\psi)$ represents the minimum mean square value of $W(x,y\,;s)$ over the region S. In (24) we can now separate the modulus and argument, giving

$$\theta(s,\psi) = ks\overline{W(s,\psi)} \quad \ldots\ldots(26)$$

$$M(s,\psi) = 1 - \frac{2\pi^2s^2}{\lambda^2}K(s,\psi) \quad \ldots\ldots(27)$$

for the phase shift and relative contrast modulation respectively.

The formulae (26) and (27) are, of course, approximate only, but we shall show that the expression for $M(s,\psi)$ may be converted to an inequality, which is valid provided the right-hand side of (27) is greater than zero. In obtaining (24) the error in the imaginary part of the integral in (23) is that implied in the approximation

$$\iint_S \sin\{ks(W-\overline{W})\}\,dS = ks\iint_S (W-\overline{W})\,dS = 0.$$

Since the mean value of $W-\overline{W}$ over S is zero, the argument of the sine will tend to be as often negative as positive. In consequence, in regions where the numerical values of $W-\overline{W}$ are in excess of the value for which the above approximation is valid, these larger values will tend to be of opposite sign, and moreover attained only over small regions of S. There will thus tend to be some cancelling of the errors occurring in such regions. If the approximation is not valid the imaginary part of the integral in (23) will have a value equal to the error itself. The relation (27) assumes this to be zero. It follows that this part of the approximation implicit in (27) will cause a negative error in the modulus $M(s,\psi)$. Again, the value used for the real part of the integral in (23) has the error implied in the approximation

$$\iint_S \cos\{ks(W-\overline{W})\}\,dS = \iint_S \left\{1-\frac{k^2s^2}{2}(W-\overline{W})^2\right\}dS = 1-\frac{2\pi^2s^2}{\lambda^2}K(s,\psi)$$

and, since $\cos\theta \geqslant 1-\frac{1}{2}\theta^2$, the error is always negative. Provided the last of these three expressions is not negative, therefore, the approximation leads to a negative error in $M(s,\psi)$. Hence, in place of (27), we may write

$$M(s,\psi) \geqslant 1-\frac{2\pi^2s^2}{\lambda^2}K(s,\psi), \qquad \ldots\ldots(28)$$

provided only that the right-hand side of (28) is greater than zero.

We propose to use the condition $M(s,\psi) \geqslant 0{\cdot}80$ as the criterion for the permissible aberration in an optical system. This criterion relates, of course, merely to a single frequency s with the orientation ψ. In view of (28), this tolerance condition may be written

$$K(s,\psi) \leqslant \frac{0{\cdot}10\lambda^2}{\pi^2s^2}, \qquad \ldots\ldots(29)$$

from which, we shall see, both the optimum form of aberration correction and the permissible aberration for the spatial frequency (s,ψ) may be found. When (29) is satisfied the contrast of the image of the component of spatial frequency s will be not less than 0·80 times the contrast which would be obtained for that frequency in the absence of aberration and in the true focal plane.

The aberration terms having odd symmetry, such as coma, give rise to terms in $W(x,y;s)$ which are independent of x and y. We shall denote these terms by $C(s)$, and write

$$W(x,y;s) = C(s) + \mathrm{E}(x,y;s) + \mathrm{O}(x,y;s) \qquad \ldots\ldots(30)$$

where $\mathrm{E}(x,y;s)$ consists of the terms in $W(x,y;s)$, other than those of $C(s)$, which are even functions of both x and y, and $\mathrm{O}(x,y;s)$ comprises the terms containing odd powers of either x or y, or of both. In the case of a circular pupil, the region of integration S is symmetrical about both the x and y axes. In this case, therefore, (22) reduces to

$$\overline{W(s,\psi)} = C(s) + \iint_S \mathrm{E}(x,y;s)\,dS \qquad \ldots\ldots(31)$$

and, from (25), we find

$$K(s,\psi)=\iint_S \{[\mathrm{O}(x,y;s)]^2+[\mathrm{E}(x,y;s)]^2\}\,dS-\left\{\iint_S \mathrm{E}(x,y;s)\,dS\right\}^2 \quad \ldots\ldots(32)$$

for the value of $K(s,\psi)$. It is only aberrations of odd symmetry, such as coma, that give rise to the terms in $\mathrm{E}(x,y;s)$. When $\psi=\frac{1}{2}\pi$ these same aberrations give rise to no terms in $\mathrm{O}(x,y;s)$. For line structures perpendicular to the meridian section of the optical system, therefore, the addition of aberration terms of even symmetry, such as spherical aberration or longitudinal shift of focus, will only further impair the contrast, since such aberrations give rise only to terms in $\mathrm{O}(x,y;s)$.

The calculation of $K(s,\psi)$ and $\overline{W(s,\psi)}$ by means of (25) and (22), or alternatively by (32) and (31), requires the evaluation of integrals of the form

$$H_{k,l}(s)=\iint_S x^k y^l\,dS, \quad \ldots\ldots(33)$$

since the integrands consist of polynomials of the form $\sum\sum g_{k,l}(s)x^k y^l$, the coefficients $g_{k,l}(s)$ involving either the aberration coefficients themselves or their squares and quadratic products. When the region S is that common to two unit circles with centres at $(\pm\frac{1}{2}s, 0)$, the integrals $H_{k,l}(s)$ may be expressed in terms of incomplete beta functions, but interpolation from the published tables to the requisite accuracy proves impossible. For this reason these integrals have been tabulated for this case for the range of values of s likely to occur in practice, and for $k+l\leqslant 12$. These tables suffice for cases where the aberration terms depend on powers of the aperture not exceeding the sixth, which will include most optical systems other than high-aperture microscope objectives. The recurrence formulae used in compiling this table readily permit the extension of the table to larger values of $k+l$. It is, however, to be noted that it is both simpler and more satisfactory to employ the Maréchal treatment (Maréchal 1948) for cases, such as microscope objectives, in which good reproduction in the image is required for all spatial frequencies.

The aberration coefficients occur in $K(s,\psi)$ as a homogeneous quadratic function of the aberration coefficients, so that the form of correction giving a maximum response for a given spatial frequency and orientation, obtained by minimizing $K(s,\psi)$, leads to unique values of the ratios of the aberration coefficients.

In the above we have considered only the case of incoherent objects. When the object is illuminated with partially coherent light, the response $C(m,p)$ for a frequency pair (m,p) is calculated in the same manner, except that the region of integration S is that common to two unit circles centred on the points $(-m,0)$ and $(-p,0)$, and an effective source (Hopkins 1953). With this interpretation on S, the treatment given above leads directly to tolerances on aberration for optical systems used in partially coherent light.

In what follows, however, we shall confine attention to image formation with incoherent light. The general method will be applied to cases of particular aberrations, the pupil being assumed circular.

When different aberration terms are present simultaneously, it is possible to find the combination of aberration coefficients which minimizes $K(s,\psi)$ and therefore gives the maximum response, $M(s,\psi)$, for a given spatial frequency. In this way both the 'balancing' of higher order aberrations by means of primary aberration terms and the choice of focal plane may be determined to give maximum contrast for any given spatial frequency. It should be noted, however, that the type of correction to which these considerations lead is not necessarily that which will give the highest limiting resolution.

§ 4. TOLERANCE ON DEFECT OF FOCUS

In the absence of true aberration, the wave front aberration with reference to a plane other than the true focal plane is of the form

$$W_0(x_0, y_0) = w_{20}(x_0^2 + y_0^2),$$

the coefficient w_{20} measuring the optical path difference introduced for the marginal ray. On a rotation of axes, $x_0^2 + y_0^2$ transforms into $x^2 + y^2$. Hence, for any angle ψ, $W(x, y) = w_{20}(x^2 + y^2)$ and, from (18), $W(x, y; s) = 2w_{20}x$, so that the various terms in (30 are

$$C(s) = \mathrm{E}(x, y; s) = 0, \qquad \mathrm{O}(x, y; s) = 2w_{20}x.$$

Substitution of these values in (32) and (31) gives

$$K(s, \psi) = 4w_{20}^2 H_{2,0}(s), \qquad \overline{W(s, \psi)} = 0$$

the notation defined in (33) being employed for the integral involved. It follows that there is no phase shift, and the tolerance for w_{20} is found by putting

$$K(s, \psi) = 4w_{10}^2 H_{2,0}(s) = \frac{0{\cdot}10\lambda^2}{\pi^2 s^2}$$

according to (29). With sufficient accuracy, therefore, the maximum permissible shift of focus, as a function of spatial frequency, is given by the formula

$$w_{20} = \pm \frac{0{\cdot}10\lambda}{2s\{H_{2,0}(s)\}^{1/2}}. \qquad \ldots\ldots(34)$$

The relative response is then not less than 0·80, provided the numerical value of w_{20} does not exceed that given by (34). The value of $D(s)$ obtained at this limit is, of course, equal to $0{\cdot}80D_0(s)$. The permissible defect of focus (34) is shown as a function of s in figure 3.

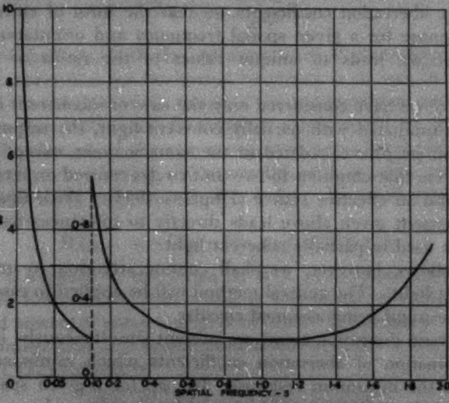


Figure 3. Tolerance for defect of focus.

For smaller frequencies, $s \leqslant 0{\cdot}10$, the value of $H_{2,0}(s)$ differs but little from 1/4, so that, in these cases, the tolerance formula (34) becomes

$$w_{20} = \pm 0{\cdot}10\lambda/S. \qquad \ldots\ldots(35)$$

Using the relations

$$w_{20} = \tfrac{1}{2}n' \sin^2 \alpha' \, \delta Z', \qquad s = \frac{\lambda}{n' \sin \alpha'} R',$$

$\delta Z'$ being the axial displacement of the focal plane, (35) leads to the form

$$\delta Z' = \pm \frac{0{\cdot}20}{R' \sin \alpha'}, \qquad \ldots\ldots(36)$$

which is the result obtained earlier for small values of s by a different method (Hopkins 1955). The error in the above formulae does not exceed 10% even for $s = 0{\cdot}20$, and is such that the values given are too small.

§ 5. ASTIGMATISM AND FIELD CURVATURE

If w_{20}, $w_{20} + w_{22}$ measure the difference in focus between the chosen image plane and the sagittal and tangential foci respectively, the aberration function of the optical system is

$$W_0(x_0, y_0) = w_{20}(x_0{}^2 + y_0{}^2) + w_{22}y_0{}^2$$

which, on rotation of axes, gives

$$W(x, y) = w_{20}(x^2 + y^2) + w_{22}(x^2 \sin^2 \psi + xy \sin 2\psi + y^2 \cos^2 \psi),$$

so that (18) becomes

$$W(x, y\,;\, s) = 2w_{20}x + 2w_{22}x \sin^2 \psi + w_{22}y \sin 2\psi$$

and the three terms introduced in (30) are

$$C(s) = \mathrm{E}(x, y\,;\, s) = 0,$$

$$\mathrm{O}(x, y\,;\, s) = 2w_{20}x + 2w_{22}x \sin^2 \psi + w_{22}y \sin 2\psi.$$

Using these expressions in (32), (31) leads to

$$\left.\begin{aligned} K(s, \psi) = 4w_{20}{}^2H_{2,0}(s) + 4w_{22}{}^2 \sin^4 \psi H_{2,0}(s) + w_{22}{}^2 \sin^2 2\psi H_{0,2}(s) \\ + 8w_{22}w_{20} \sin^2 \psi H_{2,0}(s) \\ \overline{W(s, \psi)} = 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \end{aligned}\right\} \quad \ldots\ldots(37)$$

the terms in xy disappearing after integration over the symmetrical region S.

It follows at once from (37) that astigmatism produces no spatial phase shift, but the contrast modulation $M(s, \psi) = 1 - (2\pi^2s^2/\lambda^2)K(s, \psi)$ varies with the azimuth ψ. For a line structure with direction parallel with the tangential section we put $\psi = 0$, and (37) then leads to the same tolerance as for a simple defect of focus (34), independent of the presence of astigmatism. When $\psi = \pi/2$,

$$K(s, \psi) = 4(w_{20} + w_{22})^2H_{2,0}(s)$$

and, consequently, the influence of aberrations on a line structure parallel with the sagittal section is also the same as a pure defect of focus. At the appropriate focal lines, therefore, the images of these structures are unaffected by astigmatism.

The average value of the relative contrast modulation over all azimuths is

$$\begin{aligned} \overline{M(s, \psi)} = \frac{1}{2\pi}\int_0^{2\pi} M(s, \psi)\, d\psi = 1 - \frac{2\pi^2s^2}{\lambda^2}\{4w_{20}{}^2H_{2,0}(s) \\ + \tfrac{3}{2}w_{22}{}^2H_{2,0}(s) + \tfrac{1}{2}w_{22}{}^2H_{0,2}(s) + 4w_{22}w_{20}H_{2,0}(s)\} \quad \ldots\ldots(38) \end{aligned}$$

and this has its maximum value when $w_{20}=-1/2w_{22}$. Hence, with this criterion, the best focal plane for most objects will be at the mid-point of the two focal lines. For this focal plane, the condition that the average value of $M(s)$ shall be not less than 0·80 gives the tolerance

$$w_{22}=\pm\frac{0{\cdot}10\lambda}{s}\{\tfrac{1}{2}H_{0,2}(s)+\tfrac{1}{2}H_{2,0}(s)\}^{-1/2}.$$

This formula gives the amount of astigmatism which is permissible when the image is viewed in the plane at the centre of the astigmatic range. For smaller values of s ($\leqslant 0{\cdot}10$) we obtain the approximate formula

$$w_{22}=\pm\frac{0{\cdot}20\lambda}{s} \qquad \text{......(39)}$$

since, in this case, we put $H_{0,2}(s)=H_{2,0}(s)=0{\cdot}25$. The tolerance for astigmatism (39) is twice that for defect of focus (35). This result was obtained earlier by De (1955), using a different method. In common with (35), the above expression gives values of w_{22} which are too small when $0{\cdot}10<s<0{\cdot}20$, the error being always less than 10% even for $s=0{\cdot}20$.

It has so far been assumed that we are at liberty to choose the best focal plane, but this is not always permissible. If, for example, the astigmatism and image curvature in a system are primary aberration terms, the coefficients w_{20}, w_{22} are not independent. In fact w_{20} has the form

$$w_{20}=w_{\mathrm{P}}+\tfrac{1}{2}w_{22} \qquad \text{......([illegible])}$$

where w_{P} measures the shift of focus from the gaussian image plane to the so-called Petzval surface (Hopkins 1950). In these circumstances, one is required to know what amount of astigmatism w_{22} will best compensate a given image curvature w_{P}. For such an image viewed in the gaussian focal plane, (38) becomes

$$\overline{M(s,\psi)}=1-\frac{2\pi^2s^2}{\lambda^2}\{4(w_{\mathrm{P}}+\tfrac{1}{2}w_{22})^2H_{2,0}(s)+\tfrac{3}{2}w_{22}{}^2H_{2,0}(s)$$
$$+\tfrac{1}{2}w_{22}{}^2H_{0,2}(s)+4w_{22}(w_{\mathrm{P}}+\tfrac{1}{2}w_{22})H_{2,0}(s)\}, \qquad \text{......(41)}$$

which expression is a maximum with respect to w_{22} when

$$w_{22}=-\frac{8H_{2,0}(s)}{9H_{2,0}(s)+H_{0,2}(s)}w_{\mathrm{P}}.$$

The best value of the astigmatism coefficient is therefore a function of the frequency s. For small values of s, we may write $H_{2,0}(s)=H_{0,2}(s)=0{\cdot}25$. This relation then becomes

$$w_{22}=-\tfrac{4}{5}w_{\mathrm{P}} \qquad \text{......(42)}$$

and the best form of correction is independent of frequency. The distances from the focal plane to the sagittal and tangential foci are then proportional to $w_{20}=3/5w_{\mathrm{P}}$ and $w_{20}+w_{22}=-1/5w_{\mathrm{P}}$ respectively. Hence, these two foci are on opposite sides of the focal plane, which plane divides the distance between them in the ratio 3 : 1, being nearer the tangential focus.

The image quality could be improved in this case by displacing the chosen focal plane in the direction of the Petzval surface, and allowing the response for the axial image to deteriorate slightly. This would merely require w_{P} in (40) to be regarded as the focal shift from the new image plane to the Petzval surface, and the above relations remain valid. In particular, the astigmatism coefficient w_{22} has now to be so chosen that the new image plane lies at that point of quadrisection of the inter-focal astigmatic distance which lies nearest the tangential focus.

If the focal plane is chosen at the point midway between the sagittal and tangential foci, the maximum permissible separation $\delta A'$ between these foci is

$$\delta A' = \pm \frac{0{\cdot}40}{R' \sin \alpha'}, \qquad \ldots\ldots(43)$$

R' and α' having the same significance as in (36). For example, to obtain a relative contrast modulation greater than or equal to 0·80 for $R' = 20$ lines/mm, with a system of relative aperture F/5, requires that $|\delta A'|$ be not greater than 0·20 mm.

If the value of w_{22} in (42) is substituted in (41), and the resulting expression for $\overline{M(s, \psi)}$ equated to 0·80, we obtain the expression

$$w_P = \pm 0{\cdot}22\lambda/s \qquad \ldots\ldots(44)$$

for the tolerance of the focal shift between the chosen focal plane and the Petzval surface, provided s is small. If $\delta F'$ is the distance from the gaussian focal plane to the Petzval surface, the tolerance (44) may be written

$$\delta F' = \pm \frac{0{\cdot}44\lambda}{R' \sin \alpha'}, \qquad \ldots\ldots(45)$$

exactly as in (36) above. If, therefore, the best state of correction for astigmatism is adopted relative to the gaussian image plane, the tolerance on Petzval curvature is more than double that for a pure defect of focus. If the tolerance (36) is allowed for the axial image, the permissible Petzval curvature is measured by

$$\delta F' = \pm \left(\delta Z' + \frac{0{\cdot}44\lambda}{R' \sin \alpha'} \right) = \pm \frac{0{\cdot}64\lambda}{R' \sin \alpha'}, \qquad \ldots\ldots(46)$$

it being assumed that the astigmatism is adjusted to give the best average response in the new focal plane.

§ 6. RESPONSE IN THE PRESENCE OF SPHERICAL ABERRATION

If an optical system has a higher order spherical aberration term of the form $w_{60}(x^2+y^2)^3$, the quality of the image may be improved by the introduction of an appropriate primary spherical aberration term $w_{40}(x^2+y^2)$ and by a suitable choice of focal plane, which latter will introduce a term $w_{20}(x^2+y^2)$ in the aberration function. For a structure in any azimuth the aberration function will then be of the form

$$W(x, y) = w_{20}(x^2+y^2) + w_{40}(x^2+y^2)^2 + w_{60}(x^2+y^2)^3, \qquad \ldots\ldots(47)$$

and the three groups of terms (30), which comprise $W(x, y; s)$, are given by

$$C(s) = \mathrm{E}(x, y; s) = 0,$$

$$\mathrm{O}(x, y; s) = 2xw_{20} + \{4(x^2+y^2)x + s^2x\}w_{40} + \{6(x^2+y^2)^2x + s^2(5x^2+3y^2)x + \tfrac{3}{8}s^4x\}w_{60}.$$

The mean value $\overline{W(s, \psi)} = \iint_S \mathrm{E}\, dS$ is thus zero, and there is consequently no phase shift. Substituting the above value of $O(x, y; s)$ in (32) then gives

$$\begin{aligned}K(s,\psi) = {} & [4H_{2,0}(s)]w_{20}^2 + [16H_{6,0}(s) + 32H_{4,2}(s) + 16H_{2,4}(s) + s^2\{8H_{4,0}(s) + 8H_{2,2}(s)\} \\ & + s^4\{H_{2,0}(s)\}]w_{40}^2 + [36H_{10,0}(s) + 144H_{8,2}(s) + 216H_{6,4}(s) + 144H_{4,6}(s) \\ & + 36H_{2,8}(s) + s^2\{60H_{8,0}(s) + 156H_{6,2}(s) + 132H_{4,4}(s) + 36H_{2,6}(s)\} \\ & + s^4\{29{\cdot}5H_{6,0}(s) + 39H_{4,2}(s) + 13{\cdot}5H_{2,4}(s)\} + s^6\{3{\cdot}75H_{4,0}(s) + 2{\cdot}25H_{2,2}(s)\} \\ & + s^8\{0{\cdot}1406H_{2,0}(s)\}]w_{60}^2 + [16H_{4,0}(s) + 16H_{2,2}(s) + s^2\{4H_{2,0}(s)\}]w_{20}w_{40} \\ & + [48H_{8,0}(s) + 144H_{6,2}(s) + 144H_{4,4}(s) + 48H_{2,6}(s) + s^2\{52H_{6,0}(s) \\ & + 88H_{4,2}(s) + 36H_{2,4}(s)\} + s^4\{13H_{4,0}(s) + 9H_{2,2}(s)\} + s^6\{0{\cdot}75H_{2,0}(s)\}]w_{40}w_{60} \\ & + [24H_{6,0}(s) + 48H_{4,2}(s) + 24H_{2,4}(s) + s^2\{20H_{4,0}(s) + 12H_{2,2}(s)\} \\ & + s^4\{1{\cdot}5H_{2,0}(s)\}]w_{60}w_{20} \qquad \ldots\ldots(48)\end{aligned}$$

the functions $H_{k,l}(s)$ being those defined in (33) above.

The expression (48) is of the second degree in w_{20}, w_{40}, w_{60} and values of the ratios w_{40}/w_{60}, w_{20}/w_{60} may therefore be found which maximize $M(s, \psi)$. These ratios then determine, for any given spatial frequency s, the amount of primary spherical aberration w_{40} and the shift of focus w_{20} from the gaussian image plane which will best compensate the effect of the presence of a higher order spherical aberration w_{60}. If the higher order term is so large that, for the given frequency, the relative response is significantly less than 0·80, the formulae obtained above are no longer valid, and in consequence the best form of correction and choice of focal plane may then differ from those which maximize the expression (48). The maximum permissible higher order aberration, being that for which $M(s, \psi) \geqslant 0{\cdot}80$, may, of course, be found using the ratios so determined—and equating the resulting expression in w_{60}^2 to 0·80. The dependence of the tolerance on w_{60} with s is shown graphically in figure 4 together with the necessary values of the ratios w_{40}/w_{60}, w_{20}/w_{60}.

For smaller values of s the following useful approximate expressions for the maximum permissible higher order aberration w_{60}, and the corresponding values

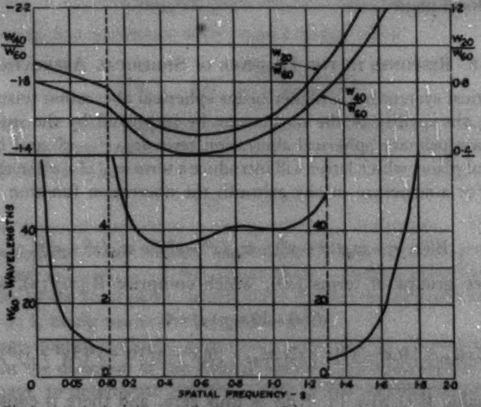


Figure 4. Optimum correction and tolerance for secondary spherical aberration.

of the ratios w_{40}/w_{60}, w_{20}/w_{60} may be obtained:

$$\left.\begin{aligned} w_{60} &= \pm\left(\frac{0\cdot 58}{s}+1\cdot 5\right)\lambda, \\ \frac{w_{40}}{w_{60}} &= -(1\cdot 80-1\cdot 5s+1\cdot 5s^2), \\ \frac{w_{20}}{w_{60}} &= +(0\cdot 90-1\cdot 4s+2\cdot 1s^2). \end{aligned}\right\} \qquad \ldots\ldots(49)$$

The error in these formulae does not exceed $0\cdot 10\,\lambda$, provided that $s<0\cdot 20$.

When there is no higher order spherical aberration present in the optical system, we put $w_{60}=0$, and the value of the ratio w_{20}/w_{40} which then maximizes (48) for any frequency determines the focal plane in which the best contrast is obtained for that frequency in the presence of primary spherical aberration. This will not necessarily be the best focal plane if the amount of primary spherical aberration is so large that the relative response $M(s)$ falls significantly below 0·80 in that plane. The tolerance on primary spherical aberration is obtained by inserting the optimal value of the ratio w_{20}/w_{40} in (48) and equating the resulting expression to 0·80. The tolerance on w_{40} so obtained is shown as a function of s in figure 5, together with corresponding values of the ratio w_{20}/w_{40}.

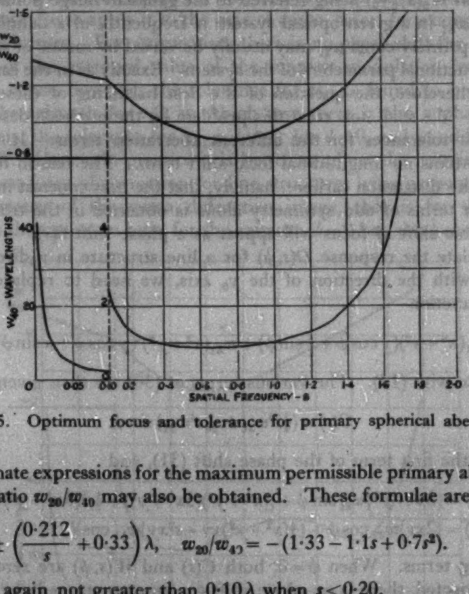


Figure 5. Optimum focus and tolerance for primary spherical aberration.

Approximate expressions for the maximum permissible primary aberration w_{40} and for the ratio w_{20}/w_{40} may also be obtained. These formulae are

$$w_{40}=\pm\left(\frac{0\cdot 212}{s}+0\cdot 33\right)\lambda, \quad w_{20}/w_{40}=-(1\cdot 33-1\cdot 1s+0\cdot 7s^2). \qquad \ldots\ldots(50)$$

The error is again not greater than $0\cdot 10\,\lambda$ when $s<0\cdot 20$.

If the optical system is of relative aperture F/5, a frequency $s=0\cdot 20$ corresponds to 40 lines mm^{-1} for $\lambda=0\cdot 5\,\mu$. The approximate formulae (49), (50) may therefore be applied over a useful practical range of frequencies. When

$s > 0{\cdot}20$ the tolerances for higher order and primary spherical aberration respectively are approximately 4λ and 1λ. These are, in fact, the tolerances obtained by Maréchal for these cases. Similarly, the tolerance on defect on focus is very nearly equal to the Rayleigh tolerance, namely $\lambda/4$, when $s < 0{\cdot}20$. For frequencies less than $s = 0{\cdot}20$, therefore, the formulae (49), (50), which give larger tolerances, may be used: when $s > 0{\cdot}20$, the Maréchal tolerances and forms of correction are more appropriate.

§ 7. TOLERANCES AND OPTIMUM CORRECTION FOR COMA

The comatic aberration terms may be classed as either circular, elliptical, or higher types of coma, according to the dependence on the azimuthal coordinate in the exit pupil. Thus, if (r, ϕ) are polar coordinates over the reference sphere, the terms of circular coma are of the form $w_{31}r^3 \cos\phi$, $w_{51}r^5 \cos\phi, \ldots, w_{n1}r^n \cos\phi$, n being a positive odd integer greater than unity. For most systems, other than high-aperture microscope objectives, the circular coma may be adequately represented by a two-term polynomial. In rectangular coordinates this has the form

$$W_0(x_0, y_0) = w_{31}(x_0^2 + y_0^2)y_0 + w_{51}(x_0^2 + y_0^2)^2 y_0, \qquad \ldots\ldots(51)$$

the aberration $W_0(x,{}_0y_0)$ being referred to the gaussian image point. The higher order coma w_{51} in a given optical system is frequently of a definite magnitude, whereas the primary coma w_{31} may usually be varied by means of simple changes in the constructional parameters of the system. Exactly as in the case of spherical aberration, therefore, the question of the best balancing of these two comatic terms arises. We shall now treat this problem by the methods described earlier, and establish tolerances for the different aberration terms. It will be noted that (51) contains no longitudinal focal-shift term. The reason for this is that outlined in the discussion earlier: namely, that the best contrast in the presence of aberration terms of odd symmetry alone is obtained in the true focal plane. Any transverse shift of focus will appear as a phase shift $\theta(s)$.

To calculate the response $D(s, \psi)$ for a line structure in a direction making an angle ψ with the direction of the y_0 axis, we need to replace (51) by the aberration function

$$W(x, y) = w_{31}(x^2 + y^2)(y \cos\psi + x \sin\psi) + w_{51}(x^2 + y^2)^2(y \cos\psi + x \sin\psi), \quad \ldots.(52)$$

in accordance with (12). The various terms of (30) are then given by

$$C(s) = (\tfrac{1}{4}s^2 w_{31} + \tfrac{1}{16}s^4 w_{51}) \sin\psi, \qquad \ldots\ldots(53)$$

which gives the first term of the phase shift (31), and

$$\begin{aligned} \mathrm{E}(x, y; s) &= \{3x^2 + y^2\}w_{31} \sin\psi + \{(x^2 + y^2)(5x^2 + y^2) + \tfrac{1}{2}s^2(5x^2 + y^2)\}w_{51} \sin\psi, \\ \mathrm{O}(x, y; s) &= \{2xy\}w_{31} \cos\psi + \{4(x^2 + y^2)xy + s^2 xy\}w_{51} \cos\psi, \end{aligned}$$

for remaining terms. When $\psi = 0$, both $C(s)$ and $\overline{W}(s, \psi)$ are zero. Hence, as would be expected, there is no phase shift for line structures lying parallel with the axis of symmetry of the comatic diffraction image. For line structures in the sagittal direction, that is perpendicular to this axis of symmetry, $\psi = \pi/2$, and consequently there will generally be a finite phase shift.

Substitution of the above expression for $\mathrm{E}(x, y; s)$ and $\mathrm{O}(x, y; s)$ in (32) and (31) gives

$$\begin{aligned} K(s,\psi) = {} & \{[4H_{2,2}(s)]w_{31}^2 + [16H_{4,2}(s) + 32H_{4,4}(s) + 16H_{2,6}(s) + s^2\{8H_{4,2}(s) \\ & + 8H_{2,4}(s)\} + s^4\{H_{2,2}(s)\}]w_{51}^2 + [16H_{4,2}(s) + 16H_{2,4}(s) \\ & + s^2\{4H_{2,2}(s)\}]w_{31}w_{51}\}\cos^2\psi + \{[9H_{4,0}(s) + 6H_{2,2}(s) + H_{0,4}(s)]w_{31}^2 \\ & + [25H_{8,0}(s) + 60H_{6,2}(s) + 46H_{4,4}(s) + 12H_{2,6}(s) + H_{0,8}(s) + s^2\{25H_{6,0}(s) \\ & + 35H_{4,2}(s) + 11H_{2,4}(s) + H_{0,6}(s)\} + s^4\{6\cdot25H_{4,0}(s) + 2\cdot5H_{2,2}(s) \\ & + 0\cdot25H_{0,4}(s)\}]w_{51}^2 + [30H_{6,0}(s) + 46H_{4,2}(s) + 18H_{2,4}(s) + 2H_{0,6}(s) \\ & + s^2\{15H_{4,0}(s) + 8H_{2,2}(s) + H_{0,4}(s)\}]w_{31}w_{51}\}\sin^2\psi - \{W_0(s)\}^2 \quad \ldots\ldots(54) \end{aligned}$$

for the determination of the tolerance, and

$$\begin{aligned} \overline{W(s,\psi)} = {} & C(s) + \{[3H_{2,0}(s) + H_{0,2}(s)]w_{31} + [5H_{4,0}(s) + 6H_{2,2}(s) + H_{0,4}(s) \\ & + s^2\{2\cdot5H_{2,0}(s) + 0\cdot5H_{0,2}(s)\}]w_{51}\}\sin\psi \quad \ldots\ldots(55) \end{aligned}$$

for the mean value (29), $C(s)$ having the value (53). By means of (53), (54) and (55), the modulus and phase of the relative response may be calculated for any azimuth and spatial frequency. The functions used above are again those defined by the integral (33).

In treating the influence of astigmatism on the response of an optical system, De (1955) averaged the response over the range $\psi = 0$ to 2π. This averaging process has also been used above for the case of astigmatism, but it would not seem so appropriate so apply it to the case of coma, since, for example, the phase shift averages to zero. For astigmatism there is no phase shift for any azimuth, and the use of the averaged value of the contrast modulation leads to useful results. For coma, however, the tolerance for w_{51} is found to be significantly smaller for the azimuth $\psi = \pi/2$, in which case there is also a phase shift. In contrast to this, the tolerance for w_{31} is larger for $\psi = 0$, and in addition there is no phase shift. Hence, in practice it is better to adjust the primary coma term to give the best form of correction for line structures lying perpendicular to the axis of symmetry of the coma diffraction image, that is for $\psi = \pi/2$. For completeness the optimum ratio for the azimuth $\psi = 0$ has also been determined,

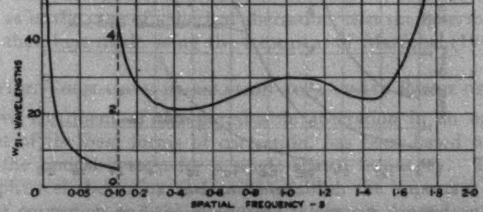


Figure 6. Optimum correction and tolerance for secondary coma ($\psi = \pi/2$).

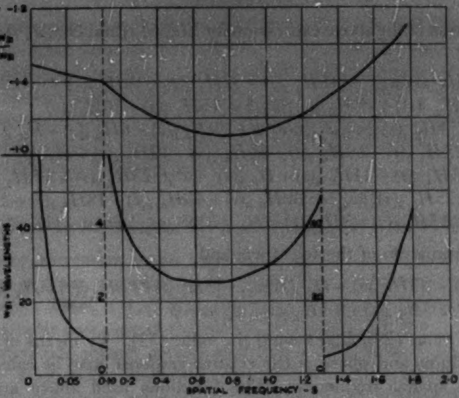


Figure 7. Optimum correction and tolerance for secondary coma ($\psi=0$).

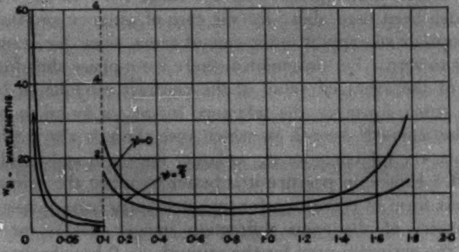


Figure 8. Aberration tolerances for primary coma.

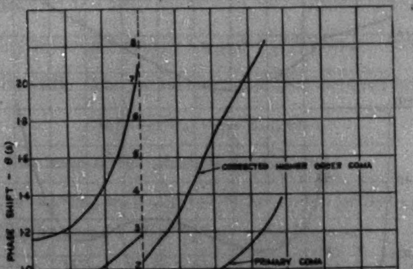


Figure 9. Phase shift $\theta(s)$ for azimuth $\psi=\pi/2$.

and it is of interest to note that for $s<0{\cdot}20$ the best form of correction is little different in the two cases.

The best forms of correction of coma for the two azimuths $\psi=0$, $\pi/2$ have been determined by minimizing the value of $K(s,\psi)$ from equation (54) with respect to w_{31}. Substitution of this value of w_{31} in (54) then gives $K(s,\psi)$ in the form of a numerical factor times $w_{51}{}^2$. The maximum permissible higher order coma, w_{51}, then follows immediately from (29). The results for the azimuths $\psi=0$, $\pi/2$ are shown graphically as functions of s in figures 6 and 7 respectively.

When $w_{51}=0$, $K(s,\psi)$ contains $w_{31}{}^2$ as a simple factor, and (29) then leads at once to the tolerances for primary coma. These results, for $\psi=0$, $\pi/2$ are shown together in figure 8.

In each of the above cases the phase shift $\theta(s)$ has been determined for the azimuth $\psi=\pi/2$. These results are shown in figure 9.

Useful approximate expressions for the maximum permissible higher order coma w_{51} and for the ratio w_{31}/w_{51} may be obtained for the range $s<0{\cdot}20$. These expressions, based on the tolerances for the azimuth $\psi=\pi/2$, are

$$w_{51}=\pm\left[\frac{0{\cdot}37}{s}+0{\cdot}83\right]\lambda,\quad w_{31}/w_{51}=-[1{\cdot}49-1{\cdot}26s+0{\cdot}85s^2],\quad\ldots\ldots(56)$$

and the phase shift is given by

$$\theta(s)=\mp[1{\cdot}16+0{\cdot}35s].\quad\ldots\ldots(57)$$

For the azimuth $\psi=0$, we find

$$w_{51}=\pm\left[\frac{0{\cdot}64}{s}+0{\cdot}96\right]\lambda,\quad w_{31}/w_{51}=-[1{\cdot}49-0{\cdot}94s+0{\cdot}41s^2],\quad\ldots\ldots(58)$$

and the phase shift is, of course, zero for this case. The corresponding results for primary coma alone are, for the azimuth $\psi=\pi/2$,

$$w_{31}=\pm\left[\frac{0{\cdot}142}{s}+0{\cdot}16\right]\lambda,\quad\ldots\ldots(59)$$

the phase shift being given by

$$\theta(s)=\mp[0{\cdot}89+0{\cdot}24s].\quad\ldots\ldots(60)$$

For the azimuth $\psi=0$, the tolerance is

$$w_{31}=\pm\left[\frac{0{\cdot}246}{s}+0{\cdot}19\right]\lambda\quad\ldots\ldots(61)$$

and the phase shift is zero. These results are applicable to the range of frequencies $s<0{\cdot}20$.

When $s>0{\cdot}20$ the above approximate expressions become less accurate: but we find, as in the case of spherical aberration, that the determination of the tolerances is then best made using the treatment of Maréchal (1948).

§ 8. ABERRATION TOLERANCES IN RELATION TO A FINITE RANGE OF FREQUENCIES

The above treatment of the permissible aberration in an optical system, and also that of the best forms of correction, has been based on the relative response of the optical system for a single spatial frequency. The numerical results for spherical aberration and coma establish the important facts that the best forms of correction and the aberration tolerances depend on which spatial frequency in the object is of interest. There is no limit to the amount of aberration allowed for the frequency $s=0$, corresponding with the simple fact that the image of an object of uniform intensity has itself a uniform distribution of intensity, no matter what the aberration of the optical system or the choice of focal plane. The permissible aberration decreases with increasing frequency, the product of the tolerance and spatial frequency remaining very roughly constant for the range $0<|s|<0{\cdot}2$. For the high frequencies, that is for s greater than about 1·5, the permissible amount of aberration increases again rapidly, and tends to infinity as the spatial frequency approaches the limit of resolution, $s=2$. For these higher frequencies, of course, even the response in the absence of aberration is very small, and consequently the difference between the response in the absence of aberration and that when $M(s,\psi)=0{\cdot}8$ becomes negligible. In this region of frequencies, therefore, the tolerance condition $M(s,\psi)\geqslant 0{\cdot}8$ has little practical significance. In practice, of course, the objects to be imaged by a given optical system usually contain a range of spatial frequencies. The frequency range which is of importance is also very often limited by the finite bandwidth of the image detector. The question then arises as to what response is obtained for lower frequencies when the aberration of an optical system is given the best form of correction and the maximum permissible aberration for a given frequency s. Consideration of the results shown graphically in figures 4, 5, 6, 7 and 8 would suggest that the response for such smaller frequencies should be, at least, not inferior to that for the given frequency, provided s is less than 0·30. Computation of a large number of cases has confirmed this, not only for the relative contrast modulation, but also for the lateral image shift $\delta u=\theta(s)/s$. Hence, the tolerance formulae given above for any given frequency s may be considered as applying to the whole band of frequencies between zero and s, it being assumed here that s is less than 0·30.

Since the lateral displacement $\delta u=\theta(s)/2\pi s$ for any component of frequency s less than s_0 (the value for which the tolerance is calculated) is found to be practically constant, the effect is to produce a lateral displacement of the image. To within an accuracy of ±5% this lateral shift may be found from the ratio $\theta(s_0)/2\pi s_0$, when the tolerable value of coma present is that for the spatial frequency s_0. A useful formula for computing the linear displacement in the image plane is

$$\delta\xi'=\frac{\theta(s_0)}{2\pi}\xi_0'$$

where ξ_0' is the length of period corresponding to the frequency s_0; or, in the notation used earlier, $\xi_0'=1/R'$.

A marked feature of the numerical results obtained earlier is the rapid decrease with increasing frequency of the aberration tolerances for the bandwidth $s\leqslant 0{\cdot}20$, and the relative constancy of the tolerances for the significant part of the frequency range above $s=0{\cdot}20$. For this latter range of frequencies, moreover, the aberration tolerances obtained here are only slightly more severe than those obtained by Maréchal using the Strehl criterion. We thus have a general rule for dealing with aberration tolerances. When $s\leqslant 0{\cdot}20$, the approximate formulae given above for the best correction and maximum permissible aberration should be employed, whereas the Maréchal tolerances are usually more appropriate when $s>0{\cdot}20$. In this latter case the quality of the optical system must be of the same order as that of a well-corrected microscope objective.

It is instructive to recall that $s=0{\cdot}20$ corresponds to 40 lines/mm for an F/5 objective and light of wavelength equal to $0{\cdot}5\,\mu$. For an objective of relative aperture F/2, $s=0{\cdot}20$ corresponds to 100 lines/mm.

We shall conclude by showing that the Strehl intensity ratio gives a measure of the mean value of the real part of the frequency response, averaged over all frequencies and azimuths. By (16) above the frequency response for a frequency s and azimuth ψ is found from the formula

$$D(s,\psi)=\frac{1}{A}\iint_{-\infty}^{+\infty} f\left(x+\frac{s}{2},y\right)f^*\left(x-\frac{s}{2},y\right)dx\,dy \qquad \ldots\ldots(62)$$

in which A is the area of the pupil, and

$$f(x,y)=f_0(x\cos\psi-y\sin\psi,\ y\cos\psi+x\sin\psi),$$

$f_0(x_0,y_0)$ denoting the pupil function with reference to fixed axes (x_0,y_0), with which the axes (x,y) make an angle ψ. Making the substitutions

$$(x\pm\tfrac{1}{2}s)\cos\psi-y\sin\psi=x_0\pm\tfrac{1}{2}s\cos\psi$$
$$y\cos\psi+(x\pm\tfrac{1}{2}s)\sin\psi=y_0\pm\tfrac{1}{2}s\sin\psi$$

we obtain, with an obvious change of notation,

$$D(s_0,\psi_0)=\frac{1}{A}\iint_{-\infty}^{+\infty} f_0(x_0+\tfrac{1}{2}s_0,y_0+\tfrac{1}{2}t_0)f_0^*(x_0-\tfrac{1}{2}s_0,y_0-\tfrac{1}{2}t_0)\,dx_0\,dy_0. \qquad \ldots\ldots(63)$$

The integral on the right-hand side may be written in terms of the transformed function $F(u',v')$, giving

$$D(s_0,\psi_0)=\frac{1}{A}\iint_{-\infty}^{+\infty}|F(u',v')|^2\exp\{-i2\pi(u's_0+v't_0)\}\,du'\,dv'$$

or, writing $G(u',v')=|F(u',v')|^2$ for the intensity in the image of a point source,

$$G(u',v')=A\iint_{-\infty}^{+\infty}D(s_0,t_0)\exp\{i2\pi(s_0u'+t_0v')\}\,ds_0\,dt_0.$$

Putting $u'=v'=0$ in this relation, we find

$$G(0,0)=A\iint_{-\infty}^{+\infty}D(s_0,t_0)\,ds_0\,dt_0$$

for the intensity at the centre of the reference sphere. In the absence of aberration the intensity in the true focal plane is given by $G_0(0,0)=A^2$, and the Strehl intensity ratio (Definitionshelligkeit) is, therefore, equal to

$$I=\frac{1}{A}\iint_{-\infty}^{+\infty}D(s_0,t_0)\,ds_0\,dt_0. \qquad \ldots\ldots(64)$$

Since $s_0=s\cos\psi$, $t_0=s\sin\psi$, the two frequencies (s,ψ), $(s,\psi+\pi)$ correspond to the frequency pairs $(\pm s_0,\pm t_0)$. According to (63), the frequency responses for the latter satisfy the relation $D(-s_0,-t_0)=D^*(s_0,t_0)$. In polar coordinates, therefore, $D(s,\psi+\pi)=D^*(s,\psi)$, where $*$ denotes a complex conjugate. In place of (64), we may now write

$$I=\frac{1}{A}\int_0^{\infty}\int_0^{\pi}\{D(s,\psi)+D^*(s,\psi)\}s\,ds\,d\psi$$

or, finally,

$$I=\frac{2}{A}\int_0^{2}\int_0^{\pi}\Re\{D(s,\psi)\}s\,ds\,d\psi \qquad \ldots\ldots(65)$$

where $\mathscr{R}$ signifies that the real part is to be taken. This is the relation stated above. The upper limit is put equal to 2, since $D(s, \psi) = 0$ for $s > 2$.

For spherical aberration, the tolerance, best form of correction, and best focal plane obtained by Maréchal (1948) are determined by the conditions

$$w_{60} = \pm 4\lambda, \qquad w_{40} = -\tfrac{3}{2} w_{60}, \qquad w_{20} = +\tfrac{3}{5} w_{60}$$

respectively. The corresponding Strehl intensity is equal to 0·80. The frequency response in this case is both real and independent of ψ, because of symmetry, so that the relation

$$I = 0{\cdot}80 = 2\int_0^2 D(s) s \, ds \qquad \ldots\ldots(66)$$

should be satisfied, a circular pupil of unit radius being assumed. Using the expression (48) the values of $M(s)$ have been calculated for this case for a range of values of s, and the product $M(s) D_0(s) s = D(s) s$ integrated by means of Simpson's rule. The right-hand side of (66) was thereby found to be equal to 0·78, a result which verifies the mutual consistency of these two treatments of aberration tolerances.

ACKNOWLEDGMENTS

I am much indebted to Miss J. M. Drewitt for checking most of the analysis and for assistance with parts of the computations, and also to Miss E. Cook, by whom the main computations were performed.

REFERENCES

DE, M., 1955, *Proc. Roy. Soc.* A, **233**, 91.

HOPKINS, H. H., 1950, *Wave Theory of Aberrations* (Oxford: Clarendon Press); 1953, *Proc. Roy. Soc.* A, **217**, 408; 1955, *Ibid.*, **231**, 91; 1956, *Proc. Phys. Soc.* B, **69**, 562.

MARÉCHAL, M., 1948, *Thesis*, University of Paris.

The Atomic Heat of Cerium between 1·5° and 20°K

By D. H. PARKINSON AND L. M. ROBERTS

Physics Department, Radar Research Establishment, Great Malvern, Worcs.

MS. received 20th December 1956

Abstract. The specific heat of a specimen of cerium of face-centred cubic structure has been measured between 1·5° and 20°K and anomalous effects, expected as a result of the susceptibility measurements of Lock, have been found. The specific heat shows a marked maximum at 12·5°K, the magnitude of which increases with the number of times the specimen has been cooled. The anomaly is probably due to antiferromagnetism.

---

§ 1. INTRODUCTION

A PRELIMINARY investigation of the atomic heats of cerium and other rare-earth metals between 1·5° and approximately 200°K was reported by Parkinson, Simon and Spedding (1951, to be referred to as PSS). Anomalous behaviour was observed in cerium between 120°K and 190°K where, with the face-centred cubic form, large hysteresis effects occurred. The work of Lawson and Ting-Yuan Tang (1949), Schuch and Sturdivant (1950) and others suggests that this is to be attributed to a $4f \rightleftharpoons 5d$ electronic transition. The atomic heat of cerium was also found to be anomalous in the range 2° to 20°K having a large hump of variable height at around 12·5°K.

Susceptibility measurements (Lock 1955, to be referred to as L) have also shown considerable thermal hysteresis between 50° and 200°K and suggest the onset of antiferromagnetism at about 12·5°K. Moreover, repeated thermal cycling of the specimen between room temperature and 20°K or less caused the thermal hysteresis loop to collapse and approximately doubled the paramagnetic susceptibility below 50°K. It was therefore thought that the variable specific heat results of PSS were due to changes in the thermal history of the samples as the experiments proceeded and, in order to elucidate this point, further specific heat measurements were undertaken. The particular object of the present work was to investigate the effect on the 12·5°K anomaly of repeated thermal cycling of the specimen between room temperature and 20°K.

§ 2. EXPERIMENTAL DETAILS

The heat capacities were measured in a Linde liquefier-cryostat, which, with a general outline of the method, has already been described (Parkinson 1954). The gold-plated copper calorimeter had two compartments and was designed specifically for work below 20°K. The heater was a 940-ohm constantan winding held with Araldite round the outside of the calorimeter. The thermometer, also on the outside, was a nominal 27-ohm Allen and Bradley carbon resistor. After its ceramic coating had been ground off, the resistor was stuck with Araldite into a thin copper sheath which was soldered to the calorimeter. Preliminary experiments showed that the thermal coupling between the calorimeter and its thermometer and heater was very good (time constant <1 sec) and several orders of magnitude better than that between the calorimeter and its surroundings (time constant $\sim 10^4$ sec).

The thermometer was calibrated by condensing liquid helium or liquid hydrogen in the upper compartment of the calorimeter and using the accepted vapour pressure data for He and H_2. Pressures were measured using a precision bore mercury manometer 15 mm in diameter. Calibration between 4·2°K and 14°K was carried out using the calorimeter itself as the bulb of a constant volume gas thermometer. Since the calibration was not exactly reproducible from one time to another, the vapour pressure portions were checked every time the calorimeter was cooled and suitable small corrections were applied to the whole range.

The specimen of cerium, which was supplied by Messrs. Johnson, Matthey and Co. weighed 12·5 g after some of the specimens used by Lock (L) had been cut from it. The chief impurities were 0·4% Li and 0·04% Fe, and x-ray analysis showed a polycrystalline face-centred cubic structure. The surface oxide was scraped off the cerium which was then weighed and placed in the lower compartment of the calorimeter. The lid was sealed on with Wood's metal and immediately the calorimeter was evacuated through a side tube, filled with He to about 2 mm pressure at room temperature and sealed off. Only a very slight dulling of the surface took place during this process.

The first sequence of measurements was made without having cooled the specimen previously in any way. The second cooling was to 20°K only, and the third cooling, to 1·5°K, was followed by a further sequence of measurements. To carry out many cycles of cooling conveniently, the specimen was removed from the calorimeter and sealed into a glass container with helium thermal exchange gas. This container was cooled bodily in liquid nitrogen and then in liquid hydrogen to 20°K, equilibrium being shown by the cessation of boiling of the liquid. It was then warmed directly to room temperature. In this way 50 cycles of cooling were accomplished in about 10 hours, after which the specimen was resealed into the calorimeter for a further sequence of measurements.

The heat capacity of the calorimeter between 1·5° and 20°K was measured with an accuracy to about 2% in a preliminary series of experiments. The ratio of the heat capacity of the specimen to that of the calorimeter was about 10 at 4·2°K, 6 at 12°K and 3·5 at 15°K. Analysis of the possible sources of error shows that individual results should be correct to 1%, except near 12·5°K, where the time for thermal equilibrium in the specimen increased, and the error could be about 2%.

§ 3. RESULTS

The experimental results are shown in figure 1. Smooth curves have been drawn through each set of points and the values of the atomic heats C_p shown in table 1 have been read from curve C, corresponding to many cycles of cooling. A small hump at about 7·0°K on the side of the main peak is distinct and outside experimental error. Since no such deviation has appeared in the measured heat capacity of the calorimeter or in the results of subsequent experiments, it is believed to be a property of the specimen. The scatter in the points for the fiftieth, fifty-first and fifty-second coolings at the peak at 12·5°K and the trough at 16°K is larger than expected and appears to be due to an intrinsic property of the specimen itself. After fifty-two coolings a further x-ray analysis showed traces of hexagonal close packed structure in addition to the original face-centred cubic form (cf. Lawson and Ting-Yuan Tang 1949).

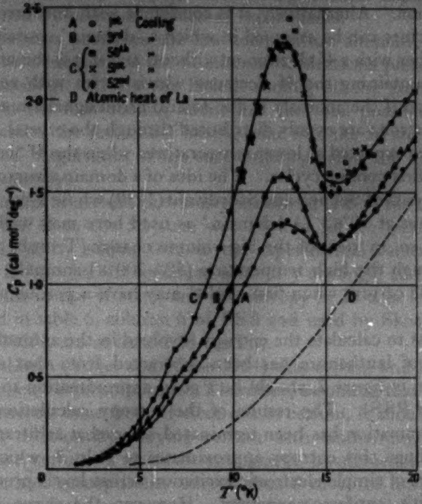


Figure 1. The atomic heat of cerium plotted against temperature.

Table 1. Atomic Heat of Cerium after 50 cycles of Cooling

T	C_p	T	C_p	T	C_p	T	C_p
1·5	0·025	5·0	0·265	9	1·10	16	1·56
2·0	0·038	5·5	0·35	10	1·45	17	1·65
2·5	0·056	6·0	0·435	11	1·85	18	1·79
3·0	0·079	6·5	0·58	12	2·21	19	1·94
3·5	0·114	7·0	0·665	13	2·31	20	2·06
4·0	0·156	7·5	0·72	14	2·00		
4·5	0·204	8·0	0·84	15	1·57		

T in °K, C_p in cal mol⁻¹ deg⁻¹

§ 4. DISCUSSION

The large thermal hysteresis in the susceptibility between 50° and 200°K (L) and similar effects in other properties of face-centred cubic cerium can be explained in terms of a $4f \rightleftharpoons 5d$ electronic transition which diminishes with successive coolings. The number of electrons remaining in the 4f state below 50°K therefore increases with each cooling, and if the specific heat anomaly at 12·5°K is caused by 4f electrons, the observed variation in size of the anomaly is also explained. The specific heat maximum coincides exactly in temperature with the maximum in the magnetic susceptibility (L) which has been tentatively ascribed to antiferromagnetism. The slightly different temperatures observed previously (PSS) may be due to impurities.

The important feature of the present results is that the peak of the specific heat anomaly stays at the same temperature after successive coolings. This is consistent with the original view (PSS) that the anomaly is due to excitation of electrons between energy levels produced by Stark splitting of the ground state of the trivalent ion. Alternatively, it is consistent with antiferromagnetism (L) if a domain structure can be assumed in which a 'domain' contains atoms of one sort only, i.e. either with a 4f electron or without, and if it is the number or size of the 'domains' containing the 4f electrons which varies with repeated cooling. On the other hand if the anomaly is due to antiferromagnetism and those atoms containing 4f electrons are evenly distributed through the crystal, the peak of the anomaly would be expected at lower temperatures when the 4f 'concentration' is lower, i.e. in the first cooling cycles. The idea of a domain structure is supported by the measurements of Schuch and Sturdivant (1950) whose X-ray studies showed two distinct phases at 90°K. A 'domain' as used here may well be a complete crystallite and if so, in view of the big volume change (Trombe and Foëx 1943) when going through the high temperature (50°–200°K) anomaly, polycrystalline face-centred cubic cerium when first cooled may have a porous structure at low temperatures.

It is of interest to calculate the entropy involved in the anomaly. To do this the atomic heat of lanthanum has been subtracted from that of cerium; the remainder, shown in figure 2, should be a good approximation to the anomalous specific heat (see PSS). The results of the entropy calculation are shown in table 2. The summation has been terminated somewhat arbitrarily at 20°K.

After 50 coolings this entropy approximates to $R \ln 2$, which is consistent both with the idea of simple electronic excitation across levels produced by Stark splitting and with antiferromagnetism. However, the shape of the curves (figure 2), with a rather sharp 'cut-off' on the high temperature side, suggests a

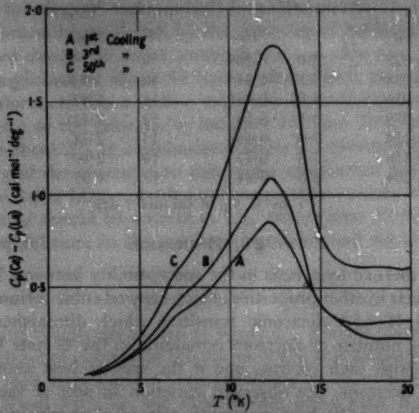


Figure 2. The atomic heat of cerium with that of lanthanum subtracted.

Table 2

No. of cooling	First	Third	Fiftieth
Anomalous entropy	0·66	0·79	1·27

$R \ln 2 = 1{\cdot}38$

co-operative phenomenon. Considering all the evidence, particularly that from the susceptibility measurements (L), antiferromagnetism is highly probable. It could be confirmed by neutron diffraction experiments.

Whichever process produces the low temperature anomaly, the anomalous entropy should be a measure of the number of 4f-electrons present per gramme atom on each cooling. If on the fiftieth cooling all the atoms contain a 4f electron, then on the first cooling about half the total possible number of 4f electrons are present, the remainder having gone into the 5d state. This agrees with the observed magnitudes of the susceptibility below the high temperature (50° to 200°K) anomaly.

In the specimens of cerium previously measured (PSS) the transfer of some electrons from the 4f to the 5d state was allowed for, in estimating the total entropy above 15°K, by adding 0·35 to the experimental values. This gave a value of 1·74 at 20°K. In our specimen after fifty coolings, all the electrons are believed to remain in the 4f state, and the entropy at 20°K is 1·90. Hence the entropies quoted in table 6, column 6 of PSS and used by Skochdopole, Griffel and Spedding (1955), are at least 0·16 too low.

It has been shown (Vogel and Klose 1954) that cerium and lanthanum form a continuous range of solid solutions. Therefore, assuming that all the magnetic phenomena of interest arise from 4f electrons in cerium, it should be possible to obtain magnetic dilution by alloying cerium with lanthanum. Then, if the behaviour in pure cerium is an antiferromagnetic phenomenon the maxima in the specific heat and susceptibility curves should move to lower temperatures and finally disappear as the concentration of cerium is reduced. An investigation of such alloys which is at present in progress may also help to explain the small hump at 7°K.

Finally, the assumption of antiferromagnetism as the source of the low-temperature anomaly does not preclude the possibility that at higher temperatures there may be further anomalies due to electronic excitation between levels produced by Stark splitting. There is a suggestion of this between 50° and 100°K in the original measurements on cerium (PSS, figure 9). Such effects should also be sensitive to the concentration of 4f electrons and hence to the thermal history of the sample as are the results described here.

ACKNOWLEDGMENTS

The authors wish to thank Miss A. E. Rennie for the X-ray analyses, and Dr. J. M. Lock for many interesting discussions. Acknowledgment is made to the Controller, H.M. Stationery Office, for permission to publish this paper.

REFERENCES

LAWSON, A. W., and TING-YUAN TANG, 1949, *Phys. Rev.*, **76**, 301.
LOCK, J. M., 1955, Conference de Physique des Basses Temperatures, Paris, p. 153.
PARKINSON, D. H., 1954, *J. Sci. Instrum.*, **31**, 178.
PARKINSON, D. H., SIMON, F. E., and SPEDDING, F. H., 1951, *Proc. Roy. Soc.* A, **207**, 137.
SCHUCH, A. F., and STURDIVANT, J. H., 1950, *J. Chem. Phys.*, **18**, 145.
SKOCHDOPOLE, R. E., GRIFFEL, M., and SPEDDING, F. H., 1955, *J. Chem. Phys.*, **23**, 2258, table III.
TROMBE, F., and FOËX, M., 1943, *C.R. Acad. Sci.* Paris, **217**, 501.
VOGEL, R., and KLOSE, H., 1954, *Z. Metallkde*, **45**, 633.

The Magnetic Susceptibility of Ytterbium from 1·3°K to 300°K

By J. M. LOCK
Radar Research Establishment, Malvern, Worcs.

MS. received 18*th January* 1957, *and in revised form* 29*th January* 1957

Abstract. The susceptibility of ytterbium metal has been measured at room temperature and between 1·3°K and 80°K. Its magnitude seems to indicate that the atoms are mostly in the ^1S state with the 4f shell completely filled, only about 1 atom in 260 being in the $^2F_{7/2}$ state with 13 electrons in the 4f shell. The paramagnetic saturation effect at the lowest temperatures seems to confirm this. The material has been tested for superconductivity down to 1·24°K in fields as low as 10 oersteds, with a negative result.

§ 1. Introduction

Ytterbium is an unusual rare-earth element both in its chemical and physical properties. For example it forms divalent as well as the usual trivalent compounds (Yost, Russell and Garner 1947), the density of the metal (6·9) is much lower than those of its neighbours thulium and lutecium (9·3 and 9·9 respectively), and its crystal structure is face-centred cubic rather than the normal hexagonal close-packed (Spedding, Daane and Herrmann 1956). These abnormal properties are associated with the tendency to form a complete 4f shell of 14 electrons in this element by the removal of one 5d electron, thus leaving only two 6s valence electrons. An independent test on the metal is provided by its magnetic susceptibility, since a completed 4f shell should be non-paramagnetic, whereas a shell with only 13 electrons should have a large magnetic moment. The susceptibility of a specimen of ytterbium has therefore been measured recently in this laboratory.

§ 2. Experimental Details

The measurement of susceptibility has been carried out by the direct force method, using a Sucksmith balance and an electromagnet capable of giving a maximum field of 11 000 oersteds at the specimen. Details of the balance and cryostat for use with liquid helium have been given elsewhere (Lock, 1957). The specimen was obtained from Johnson, Matthey and Co., Ltd., who gave the following analysis: iron 0·004%, samarium $<0·002\%$, erbium $<0·005\%$, thulium $<0·005\%$, lutecium $<0·01\%$, yttrium $<0·03\%$. It was in the form of small rough lumps of metal which had been condensed from the vapour. The first specimen, Yb1, was a single lump, measuring roughly $8 \times 6 \times \frac{1}{2}$ mm. It weighed 0·199 g and at room temperature, and in a field of 11 000 oersteds, gave a balance deflection of about 6 cm which could be measured with an accuracy of about 2%. A second specimen, Yb 2, was later made by adding three more pieces of metal giving a total weight of 0·556 g. It was used to investigate whether or not ytterbium becomes superconducting at low temperatures.

§ 3. Results

Measurements were made at room temperature and at a number of temperatures between 80°K and 1·3°K. To determine the susceptibility of the specimen, the force F on it was measured in two fields at each temperature, 11 000 oersteds and 5200 oersteds. The magnetic moment M of the specimen in these fields can then be deduced from the equation

$$M = \frac{F}{\partial H/\partial Z} \qquad \ldots\ldots(1)$$

where $\partial H/\partial Z$ is the field gradient in the vertical direction. If these values of M are plotted against H the slope of the straight line through the two points gives the susceptibility χ of the specimen, while its intercept M_0 on the axis of M gives a measure of any saturation effects, due for example to ferromagnetic impurity.

The susceptibility of Yb1 at 292°K was $0{\cdot}41 \pm 0{\cdot}01 \times 10^{-6}$ e.m.u. g^{-1} (mean of two readings). The reciprocal of its susceptibility below 80°K is plotted against

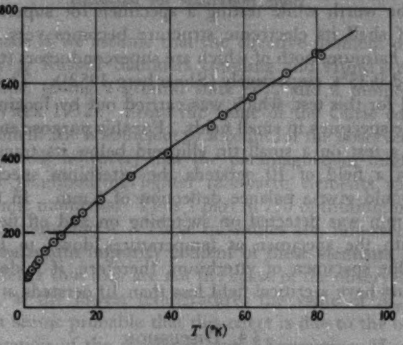


Figure 1. Yb 1. Temperature variation of $1/\chi$.

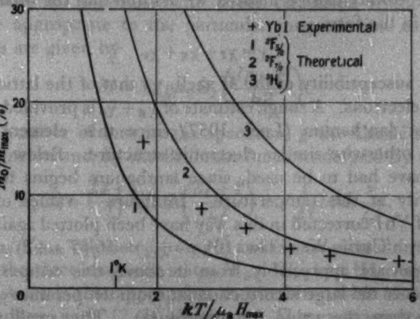


Figure 2. Saturation effect at low temperatures. The crosses denote experimental points for Yb 1.
Curves 1, 2 and 3 are the theoretical curves for $^2F_{5/2}$, $^2F_{7/2}$ and $^3H_6$ respectively.

temperature in figure 1. Above 20°K the measured values of M_0 are small and random, but below 5°K they rise appreciably to a maximum value at the lowest temperature nearly 20 times the mean high-temperature value. This is shown in figure 2 where the ordinates are values of the intercept M_0 expressed as a percentage of the magnetic moment $M_{\max}$ in the maximum field, 11 000 oersteds, and the abscissae are proportional to the absolute temperature. Because of this apparently genuine temperature-dependent saturation effect, the susceptibilities below 20°K have been calculated from the magnetic moment in the lower field only, putting in a small correction for the effect of ferromagnetic impurity, which is assumed to remain constant as the temperature is lowered. Even when calculated in this way the susceptibility will be slightly less than the true initial slope of the magnetization curve, and it is estimated that the maximum error is approximately 8% at the lowest temperature, 1·31°K, falling off rapidly as the temperature is raised (see below).

The smallness of the susceptibility of Yb1 shows the essential correctness of the idea that in the metal the 4f shell of electrons is filled. This suggested that it might be worth while testing a specimen for superconductivity, since with a filled 4f shell its electronic structure becomes very similar to that of lanthanum and hafnium, both of which are superconductors (transition temperatures 4·37°K and 0·35°K respectively (Shoenberg 1952)). The second specimen, Yb 2, was used for this test, which was carried out by looking for diamagnetic behaviour of the specimen in small fields. For this purpose an iron-free solenoid was used, and a test on a small tin ellipsoid below its transition temperature showed that in a field of 10 oersteds the ytterbium specimen, if perfectly diamagnetic, would give a balance deflection of 6 mm. In fact, no deflection greater than 1 mm was detected on switching on and off fields of 10, 17 and 35 oersteds with the specimen at temperatures down to 1·24°K, the lowest obtainable. This specimen of ytterbium, therefore, if it does become superconducting, must have a critical field less than 10 oersteds at this temperature.

§ 4. DISCUSSION

An attempt to estimate the contribution of the 4f shell of electrons to the measured susceptibility can be made if we assume that the total susceptibility χ can be written in the form

$$\chi = \chi_f + \chi_g + \chi_e \qquad \ldots\ldots(2)$$

where χ_f is the susceptibility of the 4f shell, χ_g that of the lattice and χ_e that of the conduction electrons. A rough estimate of $\chi_g + \chi_e$ is provided by the measured values of χ for lanthanum (Lock 1957) since this element has an empty 4f shell but an otherwise similar electronic structure. Below 6°K extrapolated values of χ_{La} have had to be used, since lanthanum begins to show traces of superconductivity at this temperature. In figure 3 values of the reciprocal susceptibility of Yb1 corrected in this way have been plotted against temperature. The points obey a Curie–Weiss law, $10^6(\chi - \chi_{La}) = 56/(T + 4{\cdot}2)$ up to about 50°K, but begin to deviate appreciably from it above this temperature when χ_{La} apparently becomes too large a correction (at room temperature $\chi_{La} = 0{\cdot}73 \times 10^{-6}$, actually larger than the value for ytterbium). The small Curie constant, 56×10^{-6} e.m.u. g^{-1}, indicates that, if the paramagnetism is indeed due to atoms with a 'hole' in the 4f shell, their number is very small. An estimate of this

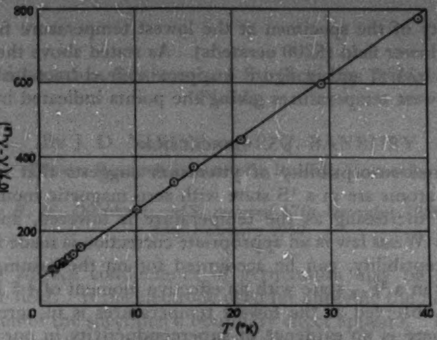


Figure 3. Yb 1. Temperature variation of $1/(\chi-\chi_{La})$. The crosses denote points corrected for saturation effect.

number can be made if we assume that the effective moment of such atoms is obtained by applying Russell–Saunders coupling and Hund's rule to an atom with 13 4f electrons, giving a ground state $^2F_{7/2}$ and a moment of 4·5 Bohr magnetons (Van Vleck 1932). From the value of the Curie constant we then find that 1 such atom in 260 is sufficient to account for the observed susceptibility.

It should be noted, however, that the paramagnetism might be due to a small trace of the neighbouring lighter rare-earth elements which have even larger effective atomic moments. Thus 1 part in 750 of thulium or 1 part in 1200 of erbium could account for it. These quantities are, however, much greater than the quoted maximum impurity content of these elements, and there is in addition other evidence that they are not the cause of the observed susceptibility. As we have seen, at the lowest temperatures the value of M_0 begins to rise appreciably, and it seems probable that this effect is due to the onset of genuine paramagnetic saturation of the magnetic ions. Expressing M_0 as a percentage of the moment $M_{\max}$ in the maximum applied field $H_{\max}$, we can compare the experimental values of $M_0/M_{\max}$ with those calculated from the theoretical saturation curve appropriate to the particular magnetic ion concerned. The saturation curves are given by

$$M = NJg\mu_B B_J\left(\frac{Jg\mu_B H}{kT}\right) \qquad \ldots\ldots(3)$$

where N is the number of ions in the specimen, g the Landé splitting factor, μ_B the Bohr magneton and B_J the Brillouin function defined by

$$B_J(x) = \frac{2J+1}{2J}\coth\left(\frac{2J+1}{2J}\right)x - \frac{1}{2J}\coth\frac{x}{2J}. \qquad \ldots\ldots(4)$$

The values of $M_0/M_{\max}$ calculated from (3) and (4) are plotted in figure 2 against $kT/\mu_B H_{\max}$ for the three magnetic states $^2F_{5/2}$, $^2F_{7/2}$ and $^3H_6$ respectively. The second of these states is that appropriate to ytterbium with 13 4f electrons and the third to thulium with 12 4f electrons. The experimental points agree reasonably with the curve for ytterbium, but not at all with that for thulium. Equations (3) and (4) also enable us to estimate the error involved in calculating the susceptibility of the specimen at the lowest temperature from its magnetic moment in the lower field (5200 oersteds). As stated above the maximum error is about 8% at 1·31°K, and in figure 3 approximate corrections have been made at the threece lowest temperatures giving the points indicated by crosses.

§ 5. CONCLUSION

The measured susceptibility of ytterbium suggests that in this metal the majority of the atoms are in a ^{1}S state with zero magnetic moment. The slight paramagnetism, increasing as the temperature is lowered, and approximately obeying a Curie–Weiss law is an appropriate correction is made for the electronic and lattice susceptibility, can be accounted for on the assumption that about 1 atom in 260 is in a $^2F_{7/2}$ state with an effective moment of 4·5 Bohr magnetons. The saturation observed at the lowest temperatures is in agreement with this hypothesis. There is no evidence of superconductivity in our specimen at the lowest temperature obtained (1·24°K).

ACKNOWLEDGEMENTS.

The author wishes to thank Dr. D. H. Parkinson for suggesting this work and Mr. J. F. C. Oliver for his assistance with the measurements. This paper is published by permission of the Controller, H.M. Stationery Office.

REFERENCES

LOCK, J. M., 1957, *Proc. Phys. Soc.* B, **70**, in the press.
SHOENBERG, D., 1952, *Superconductivity*, (Cambridge : University Press), p. 222.
SPEDDING, F. H., DAANE, A. H., and HERRMANN, K. W., 1956, *Acta Cryst., Camb.*, **9**, 559.
VAN VLECK, J. H., 1932, *The Theory of Electric and Magnetic Susceptibilities* (Oxford : Clarendon Press), p. 243.
YOST, D. M., RUSSELL, H., and GARNER, C. S., 1947, *The Rare-Earth Elements and their Compounds* (New York : John Wiley), p. 67.

Junctions Induced in Germanium Surfaces by Transverse Electric Fields

BY J. D. NIXON AND P. C. BANBURY
Physics Research Laboratory, Reading University

Communicated by R. W. Ditchburn; MS. received 17*th December* 1956, *and in final form* 13*th February* 1957

Abstract. The surface conductance of germanium may be modulated by an external electric field. If the electrode used for applying the field does not cover the full extension of the specimen a boundary effect appears. In the presence of longitudinal currents the transition regions between the field-free and field-applied surfaces behave as junctions, giving rise to carrier concentration disturbances and also to rectification. The properties of a single field-induced junction have been investigated experimentally.

§ 1. INTRODUCTION

It is well known that at the free surfaces of some semiconductors potential barriers arise as a consequence of the presence of charges trapped in surface states (Bardeen 1947). The height and thickness of these barriers can be varied by capacitively applied electrostatic fields, and the displacement of the barrier so achieved has been found to alter the values of surface conductivity and surface recombination velocity (Shockley and Pearson 1948, Henisch and Reynolds 1955, Nixon and Banbury 1956, Thomas and Rediker 1956, Many *et al.* 1956, Low 1955). In some experiments the field has been applied by an electrode not covering the full extension of the semiconductor. It has been found that under these circumstances a boundary effect may appear in the measurement of surface conductance modulation because of an additional disturbance, which will now be described.

Figure 1 illustrates the configuration of the energy bands of a strongly n-type specimen whose free surface is less n-type than the bulk (*exhaustion layer* condition),

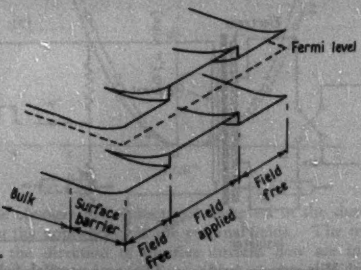


Figure 1. The energy band configuration for an exhaustion layer on germanium having a positive charge induced in the central region of the surface.

and which has a negatively charged external electrode applying a transverse field to the central portion of the specimen. There are two regions, opposite the edges of the field electrode, where there is an abrupt change in the surface barrier. These regions correspond to two junctions lying entirely within the surface. In the diagram they are of the type n^+–n^- and n^-–n^+; for other initial circumstances, p–p^+ or even p–n junctions could be produced in this way.

Under the influence of longitudinal sweeping fields within the semiconductor, such junctions might be expected to give rise to disturbances of carrier concentrations from the thermal equilibrium values, and also to rectification. Evidence has been obtained that these effects have, in some cases, been appreciable. The experiment to be described here has been carried out on an assembly which enabled the properties of a single junction to be investigated separately.

§ 2. Experimental Method

A specimen of n-type germanium, of bulk lifetime 150 μsec, resistivity 15 ohm cm, and of dimensions $2{\cdot}0 \times 1{\cdot}0 \times 0{\cdot}05$ cm was used. This specimen was fitted with soldered end contacts which were of low resistance and only slightly injecting (under the influence of a sweeping field of 10 volts cm^{-1} the conductance change due to injection was less than 0·5%). The sample, which had been etched in CP4 (Haynes and Shockley 1951), was then sandblasted along 3 mm of its length at both ends. In this condition no injection or rectification was detected; its behaviour was entirely ohmic.

Two co-planar electrodes were placed over a single large face of the specimen as indicated in figure 2. Each electrode covered half the face and extended beyond one end contact. The electrodes were insulated from the filament and from each other by polystyrene sheets, 0·009 cm thick. One electrode could be charged to a high constant positive potential and the other to an equal negative potential with respect to the germanium. The filament was placed in one arm of a Wheatstone bridge, powered from a constant current pulse generator applying sweeping fields to the germanium of about 10 volts cm^{-1}. The bridge output was displayed on a cathode ray oscilloscope and recorded photographically. The current pulses were 150 μsec long and had a repetition rate of 50 c/s. All measurements were taken with the germanium in darkness and in dry air at atmospheric pressure.

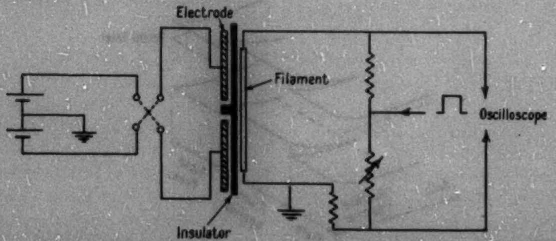


Figure 2. The pulsed Wheatstone bridge. The anti-parallel fields are applied symmetrically to the germanium by two electrodes, each covering half the filament, and extending over the end contacts.

§ 3. Experimental Results and Discussion

Upon application of the external fields the observed conductance changed. In addition it was found that the conductance varied with time in the duration of the sweeping pulse, whereas in the field-free state the conductance remained constant in this time. Since the capacitively applied fields were not switched in synchronism with the sweeping pulses it should be noted that fast transients associated with the switching of the fields (Low 1955) would not in general be observed.

All conductance changes decayed eventually to a very low value, not discernible in the drift noise of the apparatus. This latter decay corresponded to a series of time constants ranging from tens of milliseconds to several minutes. It showed the characteristic behaviour associated with the charging of 'slow' surface states; for example, once the effects had returned to zero, turning off the fields produced an effect indistinguishable from that obtained by turning on an opposite configuration from the field-free state. The specimen would then again require several minutes to attain a stable condition.

It is not possible, however, to attribute the changes occurring in the duration of a single pulse, in a time of the order of 100 μsec, to the charging of the 'slow' surface states. Nor, since they are associated with the application of the sweeping pulses rather than the switching of the capacitively applied fields, can they be related to the initial charging of the fast states.

Figure 3 shows the results obtained in the duration of the first sweeping

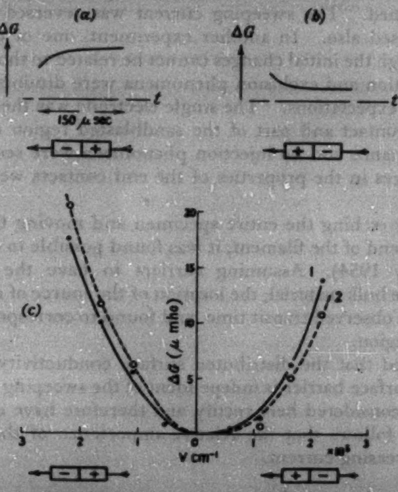


Figure 3 (*a*), (*b*). The changes in conductance observed in the duration of the sweeping pulse following the application of the external fields. The induced charge of configuration and the direction of positive current flow are indicated in each case. (*c*) Curve 1, the maximum conductance changes observed at the beginning of the sweeping pulses for the indicated configurations of induced charge. Curve 2, the maximum conductance changes observed at the ends of the sweeping pulses.

pulse after switching the capacitive fields. In each case these represent the maximum effects, which are subject to the slow decay referred to above. At the beginning of the sweeping pulse the conductance is increased from the field-free value. It is seen that these initial conductance changes are asymmetrical with respect to the reversal of the induced charge configuration, and that this asymmetry increases throughout the duration of the sweeping pulse.

The results may be interpreted in the following manner. Since the electrodes are symmetrically placed, a modulation of surface conductance which is distributed over the areas exposed to the fields must be symmetrical with respect to reversal of the induced charge configuration. This symmetry would be independent of the initial barrier height, and, if the material is homogeneous, of the magnitudes of the impressed fields, and of the distribution in energy of fast surface states. Since a junction is formed in the surface along the line of separation of the external electrodes, the observed asymmetry may be ascribed to this junction, as a rectifying action will be superimposed on the distributed change in surface conductance. The configuration shown in figure 3 (*a*) gives a junction biased in the forward direction, whereas that of figure 3 (*b*) corresponds to reverse bias. The changes in conductance occurring in the duration of the sweeping pulse are in agreement with the phenomena of carrier injection and exclusion at such junctions. In the case under consideration, the overall rectification of the system is small, since the junction is shunted by the bulk material of the filament.

Several experiments to test the self-consistency of the present interpretation have been performed. The sweeping current was reversed in direction; the asymmetries reversed also. In another experiment, one of the electrodes was removed. Although the initial changes cannot be related to those reported above, the observed injection and exclusion phenomena were diminished in magnitude in agreement with expectations. The single electrode was then moved so that it covered one end contact and part of the sandblasted region only. Very small changes in conductance and no injection phenomena were seen, indicating that possible disturbances in the properties of the end contacts were not influencing the results.

In addition, by etching the entire specimen and moving the field transition region to near one end of the filament, it was found possible to observe sweeping-out effects (Many 1954). Assuming carriers to have the normal mobility characteristic of the bulk material, the location of the source of excess carriers was estimated from the observed transit time, and found to correspond to the position of the transition region.

It may be noted that the distributed surface conductivity change resulting from a change of surface barrier is independent of the sweeping current, while the transition regions considered here rectify and therefore have current-dependent conductances. It follows that the relative importance of these junctions will diminish with decreasing current.

ACKNOWLEDGMENTS

Acknowledgment is made to Professor R. W. Ditchburn for encouragement and provision of research facilities, and also to the Admiralty for permission to publish this work.

REFERENCES

BARDEEN, J., 1947, *Phys. Rev.*, **71**, 717.
HAYNES, J. R., and SHOCKLEY, W., 1951, *Phys. Rev.*, **81**, 835.
HENISCH, H. K., and REYNOLDS, W. N., 1955, *Proc. Phys. Soc.* B, **68**, 353.
LOW, G. G. E., 1955, *Proc. Phys. Soc.* B, **68**, 10.
MANY, A., 1954, *Proc. Phys. Soc.* B, **67**, 9.
MANY, A., MARGONINSKI, Y., HARNIK, E., and ALEXANDER, E., 1956, *Phys. Rev.*, **101**, 1433.
NIXON, J. D., and BANBURY, P. C., 1956, *Proc. Phys. Soc.* B, **69**, 488.
SHOCKLEY, W., and PEARSON, G. L., 1948, *Phys. Rev.*, **74**, 232.
THOMAS, J. E., and REDIKER, R. H., 1956, *Phys. Rev.*, **101**, 984.

Fourier Images : I—The Point Source

By J. M. COWLEY and A. F. MOODIE
Division of Industrial Chemistry, Commonwealth Scientific and Industrial Research Organization, Melbourne, Australia

Communicated by A. L. G. Rees; MS. received 4th April 1956

Abstract. The problem of designing a light or electron optical system specifically for the imaging of periodic objects, such as real crystals, is discussed. A qualitative appeal to communication theory suggests that it should be possible to devise systems of higher efficiency than the conventional microscope by using the *a priori* knowledge of periodicity.

The possibility of deforming the incident wavefront in such a way that the periodic object acts as its own imaging system is considered and a formalism is set up. The particular case of the spherical wavefront is then examined in detail and it is predicted that a new type of image should be formed on certain planes. Since this image is in many ways analogous to the Fourier projection of crystallography, and since it can only be formed by periodic objects, it is named the Fourier image.

Fourier images produced experimentally with light optics are presented and shown to be in agreement with theoretical predictions. Patterns on planes other than Fourier image planes are described briefly, but detailed treatment is deferred until Part II.

The possibility of application to the imaging of crystal lattices by electron optics is discussed, but a quantitative treatment of the crucial problems of finite source size and coherence are deferred until Part III.

§ 1. Introduction

THEORETICAL and experimental enquiry into the diffraction of transverse waves tends to fall into one of two distinct categories. The first includes investigations into the diffraction of non-planar waves by general objects, and the second the diffraction of planar waves by periodic objects. Consequently two quite separate approaches have been made to problems of image-formation; that of the microscopist, and that of the structure analyst.

When attempts are made to image molecules, severe difficulties are encountered in both techniques. In the former, the problem of the design of adequate electron lenses has yet to be solved, and in the latter, a particular solution of the phase problem virtually has to be obtained for every structure. Although structure analysis is both uncertain and extremely tedious, it has so far proved to be the only method capable of yielding a detailed molecular image. Thus the success of the technique has emphasized the extreme importance of the periodic object in optics, while the complexity of its procedures has demonstrated the desirability of developing a more direct method of imaging.

A considerable literature exists on the imaging of periodic objects in optical systems capable of imaging the general object (Abbe 1873, Hopkins 1953, Berek 1950) and recently there has been extensive discussion of novel methods for imaging general objects (Gabor 1949, 1951), but no consideration has been given to the problem of designing an optical system specifically for the imaging of periodic objects.

It is the purpose of the present series of papers to discuss one approach to this problem, and to present experimental evidence to show that under suitably restricted conditions periodic objects can be imaged by means of simple optical systems which do not include lenses or mirrors as essential components.

Communication theory offers many examples of the advantages to be gained by utilizing *a priori* information associated with a given system. While the precise relationship between communication filter theory and optics is a matter for debate, the two formalisms show marked similarities, and this has prompted a number of authors to investigate the possibility of adapting an imaging system to a particular class of objects (Blanc-Lapierre 1953, MacKay 1953). With the possible exception of the design of 'super-gain antennae', this idea has not found any considerable practical application. It does, however, have an important bearing on the present problem in that conventional systems do not exploit the *a priori* knowledge of periodicity. Thus the microscopist attempts to image the periodic object in a system capable of imaging the general object and the structure analyst deduces periodicity from his diffraction patterns. It is to be expected therefore, that if this *a priori* information could be exploited, a more efficient procedure would result. A system which achieved this might, for instance, possess the same resolving power as a real, conventional system, but employ simpler apparatus, or give higher resolution with apparatus of the same quality and complexity.

It is fundamental in image-forming systems that diffracted rays must be recombined with appropriate phase and amplitude. In enquiring into systems which might exploit periodicity, it is therefore natural at the outset to enquire whether the object itself can achieve this if a suitably shaped wavefront is incident upon it. Clearly an equi-path length condition cannot be satisfied, but the possibility of the practically equivalent condition of path lengths differing by an integral number of wavelengths may exist. It is then intuitively plausible that such a condition may obtain over a plane or planes with a spherical wavefront and a plane grating.

The immediate object of this paper is to examine the possibility of imaging a plane grating by means of a point source and to compare predictions with the relevant experimental results in light optics. Discussion of the limitations inherent to a real system, that is, of effectively finite sources, coherence and aberrations in any subsidiary optics, is deferred until later publications (Cowley and Moodie 1957 a, b).

As can be appreciated from the following argument, the limitation to plane gratings need not, of itself, constitute a practical deterrent. An approximate statement of the condition that a real crystal should behave as a plane grating to a given wavelength is that the Ewald sphere should be substantially planar over a sufficiently large area in reciprocal space. This is not generally satisfied for x-rays of the wavelengths most commonly employed in structure determination, but is satisfied over a wide range of conditions for electrons in the energy ranges most frequently employed in microscopes and diffraction cameras. In other words, for a given setting of the crystal in the beam electrons of wavelengths in the region of 0·05 Å will produce a diffraction pattern which includes all of the orders associated with an appropriate zone-axis, while x-rays of wavelength in the region of 1 Å will, in general, produce no pattern at all. Direct confirmation that sufficient structural information is available at one setting of the crystal to allow of precise determination of atomic positions has been obtained by one of the authors (Cowley 1953 a, b).

§ 2. Formal Statement of the Problem

The coordinate system of figure 1 is set up. In order to simplify the analysis the axes of the unit cell are chosen to be orthogonal. The transmission function of the general plane periodic object is then defined in the usual way by

$$\rho(\xi,\eta)=\sum_h\sum_k F_{hk}\cos 2\pi\left(h\frac{\xi}{a}+k\frac{\eta}{b}+\phi_{hk}\right) \qquad \ldots\ldots(2.1)$$

where (ξ,η) are direct lattice coordinates, (h, k) are reciprocal lattice coordinates and ϕ_{hk} are the phases associated with the structure amplitudes F_{hk}.

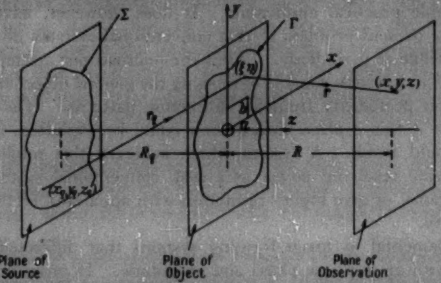


Figure 1. Coordinate system adopted in formal statement of the problem.

Suppose that the source lies on a plane parallel to the object plane and at a distance R_q from it, that (x_q, y_q, z_q) is a typical point on it, and that $\mathbf{r}_q$ is the vector from (x_q, y_q, z_q) to the point (ξ,η) on the object plane. Similarly, suppose that the plane of observation lies at a distance R from the object plane, that (x, y, z) is a point on it, and that $\mathbf{r}$ is the vector from (x, y, z) to (ξ,η).

If the source consists entirely of the point (x_q, y_q, z_q), and if the object consists of a screen with aperture Γ, then the wave function at (x, y, z) is given by the Kirchhoff integral,

$$v=\frac{1}{4\pi}\int_\Gamma (rr_q)^{-1}\left[\left(\frac{1}{r}+ik\right)\cos(\mathbf{n}\,.\,\mathbf{r})-\left(\frac{1}{r_q}+ik\right)\cos(\mathbf{n}\,.\,\mathbf{r}_q)\right]\exp\{-ik(r+r_q)\}\,dS \qquad \ldots\ldots(2.2)$$

where dS is a surface element, $\mathbf{n}$ is the unit normal and $k=2\pi/\lambda$.

The screen is now replaced by the periodic object $\rho(\xi,\eta)$ with bounds Γ and the source is described by a function $q(x_q, y_q)$ with bounds Σ. Then if dS_q is an element of surface on the source, the wave function at (x, y, z) is given by

$$u=\frac{1}{4\pi}\int_\Gamma\int_\Sigma \rho(\xi,\eta)\,q(x_q,y_q)\,(rr_q)^{-1}\left[\left(\frac{1}{r}+ik\right)\cos(\mathbf{n}\,.\,\mathbf{r})-\left(\frac{1}{r_q}+ik\right)\cos(\mathbf{n}\,.\,\mathbf{r}_q)\right]$$
$$\times\exp\{-ik(r+r_q)\}\,dS\,dS_q \qquad \ldots\ldots(2.3)$$

For any source function, the evaluation of equation (2.3) is a formidable problem. Fortunately, under the experimental conditions envisaged, extensive simplification is possible with negligible loss in accuracy. Thus both r and r_q will always be much greater than the wavelength and the maximum effective linear dimension of the object. Under these conditions (2.3) reduces to

$$u=\frac{i[\cos(\mathbf{n}\,.\,\mathbf{r})-\cos(\mathbf{n}\,.\,\mathbf{r}_q)]}{2\lambda r r_q}\int_\Gamma\int_\Sigma \rho(\xi,\eta)\,q(x_q,y_q)\exp\{-ik(r+r_q)\}\,dS\,dS_q. \qquad \ldots\ldots(2.4)$$

Under equivalent conditions the approximation is better for electrons and a real crystal than for light and a true two-dimensional grating. Considerable use will be made of this fact since many steps in the analysis are purely formal and would be difficult to justify rigorously. The main justification will be sought in the quantitative comparison of theory with experiment. It will emerge that simple quantitative light optical tests are available for each main result, so that the experimentally much more involved electron-optical tests, in which the main practical interest lies, may be deferred with confidence.

§ 3. THE POINT SOURCE AND THE INFINITE PLANE PERIODIC OBJECT

The analysis will be made in one dimension. No new feature and no additional difficulty is encountered in two dimensions; the notation merely becomes more complicated. The results for two dimensions will be quoted at the end of the section.

A conventional parabolic approximation is made to $\exp\{-ik(r+r_q)\}$ so that

$$\exp\{-ik(r+r_q)\}=\exp\{-ik(b_0+b_1\xi+b_2\xi^2)\} \qquad \ldots\ldots(3.1)$$

where $$b_0=\tfrac{1}{2}\left(\frac{x_q^2}{R_q}+\frac{x^2}{R}\right);\; b_1=-\left(\frac{x_q}{R_q}+\frac{x}{R}\right);\; b_2=\tfrac{1}{2}\left(\frac{1}{R_q}+\frac{1}{R}\right).$$

Equation (2.4) then becomes

$$u=K_1\int_{-\infty}^{\infty}\sum_h F_h\cos 2\pi\left(h\frac{\xi}{a}+\phi_h\right)\exp\{-ik(b_0+b_1\xi+b_2\xi^2)\}\,d\xi. \qquad \ldots\ldots(3.2)$$

Reversing the order of integration and summation and then performing the integration,

$$u=K_1(1-i)\left(\frac{\pi}{2kb_2}\right)^{1/2}\exp\left\{ik\left(b_0-\frac{b_1^2}{4b_2}\right)\right\}\sum_h\exp\{i\gamma_h^2\}F_h\cos 2\pi\left(\frac{h}{a}\,\frac{b_1}{b_2}-\phi_h\right) \qquad \ldots\ldots(3.3)$$

where $$\gamma_h^2=\pi\lambda\frac{R+R_q}{RR_q}\,\frac{h^2}{a^2}.$$

The dependence of (3.3) on x_q is most conveniently displayed by rewriting it in the form

$$u=K_1(1-i)\left(\frac{\pi}{2kb_2}\right)^{1/2}\exp\{ik(c_0+c_1x_q+c_2x_q^2)\}\sum_h\exp\{i\gamma_h^2\}F_h\cos 2\pi$$
$$\times\left(E_hx_q+\frac{\psi_h}{2\pi}\right) \qquad \ldots\ldots(3.4)$$

where $$c_0=R+R_q+\tfrac{1}{2}\frac{x^2}{R+R_q};\quad c_1=-\frac{x}{R+R_q};\quad c_2=\frac{1}{2(R+R_q)};$$

$$E_h=\frac{h}{a}\,\frac{R}{R+R_q};\; \psi_h=2\pi\left(\frac{R_q}{R+R_q}\,\frac{hx}{a}+\phi_h\right).$$

If the point source lies on the axis, i.e. $x_q=0$, and if constant terms are gathered under K_2^2 the intensity at a point on the plane of observation is given by

$$I=uu^*=K_2^2[\sum_h F_h^2\cos^2\psi_h+\sum_i\sum_j\cos(\gamma_i^2-\gamma_j^2)\,F_iF_j\cos\psi_i\cos\psi_j. \qquad \ldots\ldots(3.5)$$

Suppose that it is possible to satisfy the set of conditions

$$\cos(\gamma_1^2-\gamma_2^2)=\cos(\gamma_1^2-\gamma_3^2)=\ldots\ldots=\cos(\gamma_i^2-\gamma_j^2)=\ldots\ldots=+1. \qquad \ldots\ldots(3.6)$$

Equation (3.5) then represents the intensity due to a distribution

$$\rho(x/M)=K_2\sum_h F_h\cos\psi_h.$$

This distribution is identical with that of the original periodic object, but magnified by a factor

$$M=(R+R_q)/R_q. \qquad \ldots\ldots(3.7)$$

This is the magnification that would be expected from a shadow image.

The conditions (3.6) in fact define a set of image planes. These conditions are equivalent to

$$\tfrac{1}{2}\frac{\lambda}{a^2}\,\frac{RR_q}{R+R_q}=\frac{m_{12}}{1^2-2^2}=\frac{m_{13}}{1^2-3^2}=\ldots\ldots=\frac{m_{ij}}{h_i^2-h_j^2}=\ldots\ldots$$

where m_{ij} are a set of integers; or

$$\frac{1}{R}+\frac{1}{R_q}=\frac{1}{\nu}\,\frac{1}{2}\,\frac{\lambda}{a^2} \qquad \ldots\ldots(3.8)$$

where ν is an integer, possibly negative. The system thus possesses multiple focal lengths $\nu(2a^2/\lambda)$.

Conditions equivalent to (3.8) in two dimensions are

$$\tfrac{1}{2}\lambda\frac{RR_q}{R+R_q}=\ldots\ldots=m_{ij}\left[\frac{(h_i^2-h_j^2)}{a^2}+\frac{(k_i^2-k_j^2)}{b^2}\right]^{-1}. \qquad \ldots\ldots(3.9)$$

These conditions reduce to (3.8) provided

$$a/b=(n_1/n_2)^{1/2} \qquad \ldots\ldots(3.10)$$

where n_1 and n_2 are integers.

The quantitative treatment of the patterns lying between the focal planes and the comparison of the theoretical and experimental results is somewhat lengthy. Further, the best approach differs considerably from that presented above. It is therefore proposed to defer this discussion to Part II of the series. For the purposes of the next section it is, however, necessary to consider one special case, that defined by

$$\frac{1}{R}+\frac{1}{R_q}=\frac{1}{\nu+\frac{1}{2}}\;\frac{1}{2}\,\frac{\lambda}{a^2}.$$

For such planes equation (3.5) becomes

$$I=K_2^2\left[\sum_h F_h^2\cos^2\psi_h-\sum_i\sum_j F_iF_j\cos\psi_i\cos\psi_j\right].$$

This pattern is seen to result from the distribution

$$\rho\left(\frac{x}{M}\right)=\sum_h(-)^{|h|}F_h\cos 2\pi h\frac{x}{M}=\sum_h F_h\cos 2\pi h\frac{(x+\frac{1}{2})}{M}, \qquad \ldots\ldots(3.11)$$

that is, from a distribution related to the original by a magnification and a translation of half the unit-cell edge.

Thus, focal planes exist whenever

$$\frac{1}{R}+\frac{1}{R_q}=\frac{1}{\nu}\frac{1}{2}\frac{\lambda}{a^2} \qquad \ldots\ldots(3.12)$$

ν being integral or half integral, positive or negative.

In order to distinguish between the conventional image and the patterns on the planes defined by (3.12) it is proposed to describe the former as images and the latter as Fourier images. The reasons for the choice of the word Fourier will appear later. It is further proposed to adopt the notation $I(\mu)$ for the pattern on the plane defined by (3.12) and with ν replaced by μ. Thus $I(3)$ refers to the sixth real Fourier image, and $I(8{\cdot}06)$ refers to the pattern on the plane defined by

$$\frac{1}{R}+\frac{1}{R_q}=\frac{1}{8{\cdot}06}\,\frac{1}{2}\,\frac{\lambda}{a^2}.$$

When all structure amplitudes except one are zero the conditions for focus disappear, so that a cosinusoidal grating is in focus for all R. The result for this special case has been derived by Gabor (1949) in the course of his investigations into the imaging of general objects.

Hopkins (1953), in considering the imaging of periodic objects by a lens, treated what was, in effect, the case of the gratings $\cos\theta$ and $1+\cos\theta$. For the latter it was found that multiple focal planes existed in the image space of the lens.

§ 4. EXPERIMENTAL

4.1. *Qualitative Observations*

With some care the phenomenon may be observed with a fine gauze and a distant light source. It is much simpler, however, to place the object on the stage of a microscope and use a pinhole and the substage condenser to form the source. On racking the objective up and down, a series of magnified and demagnified Fourier images are observed. The normal image plane, $I(0)$, may be identified on noting that only here do non-periodic objects such as specks of dust or flaws in the grating come into sharp focus. This 'filter-like' action is discussed in greater detail in Part II.

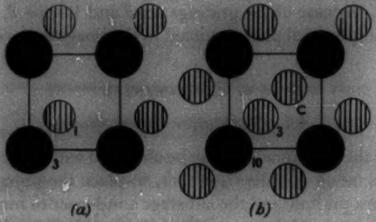
The origin of the focusing and magnification effects can be observed physically in the formation and interaction of Fresnel fringe patterns. On racking up the objective from its normal position, fringes are observed to run out from the detail in the image. Those originating from the periodic object reinforce to produce patterns of varying complexity. Finally they reinforce completely and a magnified image of the grating is formed. The cycle is repeated with increasing magnification and without loss of resolution. Beyond a certain point it is impossible to obtain a sharply focused image unless the crossover distance is set to satisfy the condition for focus at infinity, that is, $R_q=2\nu a^2/\lambda$.

On racking the objective down from its normal position, a similar sequence is observed except that the Fourier images are virtual and reduced. The process begins to fail when the orders of diffraction fail to overlap completely, and terminates on the plane of the diffraction pattern. Here the objective is focused on the source.

The origin of the filter effect can be observed physically by repeating the above procedure, but this time confining attention to the non-periodic detail. Here again fringes run out from the objects on racking away from the normal position, but now, finding no corresponding fringes to interact with, they rapidly dissipate themselves, forming a background to the periodic component.

If the grating consists simply of two orthogonal sets of ruled lines, or of a gauze, these purely qualitative observations are not completely convincing. It has become almost conventional in investigations of this type to find that the Patterson function, or some closely related hybrid, is generated, and the reasons for this are usually elusive. Since the Patterson function of a grid, or a lattice consisting of a disc on the origin of the unit cell, bears a marked resemblance to the intensity distribution in the image itself, it is desirable to employ some more distinctive array for qualitative observations.

The authors are indebted to Mr. J. L. Farrant for making available to them a variety of gratings prepared by him in the course of a series of researches undertaken by Dr. A. L. G. Rees and himself on certain diffraction effects observed in the electron microscope. A sketch of the unit cell of one of these gratings is given in figure 2 (*a*), and a sketch of the essential features of the Patterson function of this cell in figure 2 (*b*). It will be noted that the presence or absence of peak C constitutes a test of whether a Patterson function or an image has been formed. Under conditions which will be described in § (4.2), the photographs of figure 3 (*a*), (*b*) and (*c*) (Plate) were obtained. Actual values used in the experiment were, $a = 35{\cdot}3\ \mu$, $R_q = 0{\cdot}545$ cm, $R = 0{\cdot}395$ cm, and $\lambda = 5461$ Å. Figure 3 (*b*) is the normal magnified image and figure 3 (*c*) is the Fourier image, $I(\frac{1}{2})$. It will be observed that, apart from a difference in magnification and a reduction in the resolution of non-periodic detail, the plates are identical. In particular, there is no trace of peak C. Qualitatively, then, at least, the image-like properties of these patterns predominate.



(*a*) (*b*)

Figure 2. (*a*) Sketch of the unit cell of the grating employed to obtain figure 3. (*b*) Patterson function of (*a*). Approximate weights are indicated beside each peak.

4.2. *Apparatus*

In order that the source to object distance may be readily varied, it is convenient to employ the reduced image of a pinhole as a source. Under these conditions Fourier image space can be divided into four distinct regions. This is illustrated in figure 4 for a divergent pencil incident on the object. In region I the Fourier images are real and magnified, in region II they are virtual and reduced, in region III they are virtual and reduced and in region IV they are virtual and magnified.

A corresponding set of regions exists for a convergent pencil incident on the object.

In order that all of these regions might be conveniently explored the experimental arrangement shown schematically in figure 5 was employed. A 60 watt mercury discharge lamp, a pinhole of approximately 1000 μ diameter, and a No. 77 Wratten filter formed the illuminating system for a Zeiss Opton microscope. Measurements along the optical axis could then be made with considerable precision by means of the calibrated fine focus control.

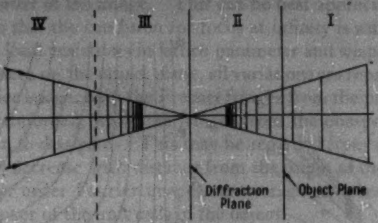


Figure 4. Distribution of Fourier image planes for a divergent pencil incident on the object.

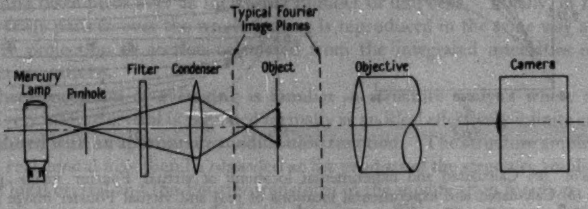


Figure 5. Experimental arrangement.

An effective source diameter of 30 μ was obtained with this system. While it might be considered that such an illuminating system by no means approximates to a point source when located only a few tenths of a millimetre from a grating of only 15 μ period, it was found that reduction of the diameter of the pinhole led to no change in the results within the limits of experimental error, but merely increased exposure times. The value given above was therefore adopted for experimental convenience. A quantitative discussion of this point, which turns out to be fundamental in the theory, is given, along with further experimental evidence, in Part III of this series.

Under these conditions exposures of approximately thirty minutes gave adequate density on Kodak Panatomic-X plates.

4.3. *Testing of Expressions for Location of Focal Planes and for Magnification*

In the experiments described in this section, a photograph of the reduced image of a gauze was employed as the periodic object. The length of the edge of the unit cell was found to be $14{\cdot}67\,\mu \pm 0{\cdot}04\,\mu$ in the mean.

The distances R_q and R were found by focusing successively on the diffraction pattern, the object, and the relevant Fourier image plane. Measurements were made for a range of values of R_q by varying the setting of the substage condenser.

The results of the measurements for both real and virtual images are compared with the calculated values in figure 6 (*a*) and (*b*). Theoretical and experimental values of the location of and the magnification on the set of Fourier image planes for which $R_q = 0{\cdot}226$ cm are compared in the table. Due to the limited traverse of the fine focus control, accuracy of measurement is low in this set. Nevertheless it can be seen that there is satisfactory agreement within the limits of experimental accuracy.

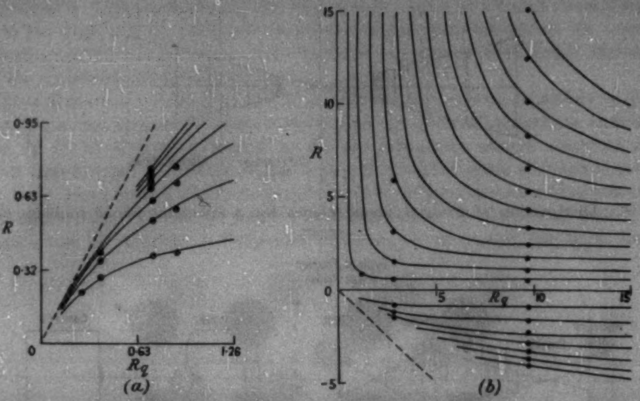


Figure 6. (*a*) Calculated and experimental locations of virtual Fourier image planes. (*b*) Calculated and experimental locations of real and virtual Fourier image planes for greater R_q. R and R_q are measured in units of $2a^2/\lambda$.

Comparison of the Theoretical and Experimental Values of the Location of and Magnification on a Series of Fourier Image Planes. $R_q = 0{\cdot}226$ cm, $a = 14{\cdot}67\,\mu$, $\lambda = 5461$ Å

	R_{calc} (cm)	R_{exp} (cm)	M_{calc}	M_{exp}
−½	−0·034	−0·033		
−1	−0·058	−0·056		
−1½	−0·078	−0·075		
−2	−0·093	−0·090		
−2½	−0·105	−0·103		
−3	−0·116	−0·115		
−3½	−0·124	−0·126		
½	0·048	0·048	1·21	1·20
1	0·121	0·119	1·52	1·51
1½	0·248	0·250	2·11	1·99
2	0·52	0·46	3·04	2·89
2½	1·54	1·13	6·00	5·70

A number of the patterns observed with this value of R_q, and reproduced in figure 7 (Plate), illustrate some of the reasons for choosing the term Fourier image.

It is clear from figure 7 (c), the object pattern, that the object deviates appreciably from ideal periodicity. Thus there is considerable local variation both in lattice parameter and in the shape of the constituent elements, together with a small regular increase in lattice parameter from the centre outwards, due to defects in the optical system employed in producing the grating.

All of these features will be reproduced in a conventional imaging system with an accuracy depending on the degree of perfection of the constituent optical components. The position is entirely different for Fourier images, and further, changes with the order of the image. This can be best appreciated by imagining R_q to be chosen so that the condition for focus at infinity is satisfied and by confining attention to local variations in lattice parameter and shape.

In $I(0)$, the pattern on the object plane, all variations are reproduced. In $I(\frac{1}{2})$, the first real Fourier image, only the Fresnel fringes from the periodic component have completely reinforced, while the fringes from the non-periodic component have already begun to dissipate. This may be regarded from a different point of view. The $|F_{hk}|$ decrease with distance from the origin of the reciprocal lattice. Hence in the lower order Fourier images, an averaging process takes place over only a limited number of the unit cells in the object.

As the order number of the Fourier image increases Fresnel fringes originating from the non-periodic component are increasingly dissipated, or alternatively, the averaging takes place over an increasing number of unit cells. Finally, in $I(\infty)$, the average pattern over the whole grating is reproduced in the same way as in a Fourier projection or section calculated from the integrated intensities of the diffraction pattern.

This conception of averaging is familiar in structure analysis where a real effectively infinite crystal is regarded formally as an ideal effectively infinite crystal convoluted with an appropriate modification function. The structure amplitudes of the real crystal may then be regarded as the product of the structure amplitudes of the ideal crystal and the Fourier transform of the modification function. It is usual to treat temperature effects in this way. Thus the pattern shown in figure 7 (a) might be described as one having a high temperature factor, since the intensity falls off rapidly with increase in reciprocal lattice coordinate. It would therefore be expected that $I(\infty)$ would exhibit relatively poor definition in this particular case, certainly considerably poorer than $I(0)$. The expected loss in definition is however seen to follow entirely from deficiencies in the object and not in the system. In fact, under these circumstances, loss in definition might be taken as a measure of the efficiency of the system.

On examining figure 7 it is, in fact, observable that the Fourier images approach the mean with increasing order number.

An isolated object constitutes an extreme case of this local variation. As can be seen such an object is sharply focused in $I(0)$, but has virtually disappeared in $I(2)$.

§ 5. DISCUSSION

The most obvious aspects of the limited theory presented here seem to be well substantiated experimentally for light optics.

The principal interest, however, lies in electron optics, where on a first inspection it would appear that even better agreement could be expected. Thus, for practical orders of magnitude the approximation of equation (2.4) holds more accurately for electrons than for photons. Further, electron sources of much greater intensity than light sources are readily available, and the limiting factor in resolution appears to be the perfection of the object. This latter is much lower for an optical grating than for most crystalline substances.

The effects of finite source size and the closely related subject of coherence, however, emerge as the crucial factors in considerations of application to practical electron-optical systems. These subjects are considered quantitatively in Parts II and III of this series and further discussion of application is therefore deferred to those papers.

References

Abbe, E., 1873, *Arch. mikrosk. Anat.*, **9**, 431.

Berek, M., 1950, *Optik*, **6**, 1.

Blanc-Lapierre, A., 1953, *Communication Theory*, Edited by Willis Jackson (London: Butterworths Scientific Publications).

Cowley, J. M., 1953 a, *Acta Cryst., Camb.*, **6**, 516; 1953 b, *Ibid.*, **6**, 522.

Cowley, J. M., and Moodie, A. F., 1957 a, *Proc. Phys. Soc.* B, **70**, 497; 1957 b, *Ibid.*, **70**, 505.

Gabor, D., 1949, *Proc. Roy. Soc.* A, **197**, 454; 1951, *Proc. Phys. Soc.* B, **64**, 449.

Hopkins, H. H., 1953, *Proc. Roy. Soc.* A, **217**, 408.

MacKay, D. M., 1953, *Communication Theory*, Edited by Willis Jackson, p. 521 (London: Butterworths Scientific Publications).

Fourier Images: II—The Out-of-focus Patterns

By J. M. COWLEY and A. F. MOODIE

Division of Industrial Chemistry, Commonwealth Scientific and Industrial Research Organization, Melbourne, Australia

Communicated by A. L. G. Rees; MS. received 4th April 1956

Abstract. The theoretical and experimental investigation of Fourier images described in Part I is extended to the patterns on planes other than focal planes.

The theory is developed from a new statement of the diffraction equations correct to the Fresnel approximations, and experimentally observed patterns are shown to be in good agreement with the results of the theory.

It is concluded that, at least within the angular range of the experiments, some fifteen degrees, the Fresnel approximation gives an adequate description of the phenomenon.

---

§ 1. Introduction

In Part I of this series (Cowley and Moodie 1957 a) the diffracting system consisting of an arbitrary coherent source and a periodic object was investigated in outline. The special case of a point source and an infinite periodic object was then investigated in detail and it was shown, theoretically and experimentally, that Fourier images were formed on a certain set of planes.

As a preliminary to the discussion of more complex systems it is desirable to discover whether the theory gives an adequate description of out-of-focus patterns, that is, patterns formed on planes lying between Fourier image planes. While it is possible to achieve this by the methods of Part I, a clumsiness in the analysis becomes apparent. Extension of the theory beyond this point soon leads to lengthy manipulation of complex functions and the underlying physical processes are obscured.

Evidently what is lacking is a model for Fresnel diffraction analogous to that currently available for Fraunhofer diffraction. Thus the conception of the reciprocal lattice, the Fourier transform and the Ewald sphere makes possible the rapid qualitative assessment of the important factors in an experiment involving plane waves so that, even when rigorous solutions cannot be obtained, it is almost invariably possible to arrive at a sufficiently accurate estimate by straightforward means. With some approximations, an attempt has been made to set up such a model, the details of the analysis being given in another publication (Cowley and Moodie, to be published). The purpose of the present communication is to describe the out-of-focus patterns observed with a point source and a periodic object in terms of this model and to compare the results with the relevant experimental evidence.

§ 2. Theoretical

The following additions are made to the notation of Part I:

$$\mathscr{F}_\alpha g(x) = \int_{-\infty}^{\infty} g(x) \exp\{-i\alpha x\}\, dx = G(\alpha)$$

so that

$$\mathscr{F}_{h,k}\rho(\xi,\eta)=F(h,k)\,; \qquad \mathscr{F}_{\alpha,\beta}\,q(x_q,y_q)=Q(\alpha,\beta)$$

$$g(x)\star g_1(x)\star\ldots\star g_n(x)=\int_{-\infty}^{\infty}g_n(u_n)\,du_n\int_{-\infty}^{\infty}g_{n-1}(u_{n-1})\,du_{n-1}\ldots$$
$$\times\int_{-\infty}^{\infty}g_1(u_1)g(x-u_1\ldots-u_n)\,du_1.$$

To the approximation of equation (3.2) of Part I, but for an arbitrary non-periodic object, it is found that the wave function can be expressed in four symmetric and equivalent forms (Cowley and Moodie 1957 c). The two of immediate interest are, apart from constants,

$$u=\exp\left\{-ik\frac{x^2}{2R}\right\}\left[F\left(-k\frac{x}{R}\right)\star\exp\left\{ik\frac{R_q}{2R(R+R_q)}x^2\right\}\star q\left(-\frac{R_q}{R}x\right)\right.$$
$$\left.\times\exp\left\{-ik\frac{R_q}{2R^2}x^2\right\}\right] \qquad \ldots\ldots(2.1)$$

$$u=\exp\left\{-ik\frac{x^2}{2R}\right\}\left[\left(\exp\left\{ik\frac{R_q}{2R(R+R_q)}x^2\right\}\rho\left(\frac{R_q}{R+R_q}x\right)\star\right.\right.$$
$$\left.\left.\times\exp\left\{-ik\frac{R_q}{2R(R+R_q)}x^2\right\}\right)\star q\left(-\frac{R_q}{R}x\right)\exp\left\{-ik\frac{R_q}{2R^2}x^2\right\}\right]. \qquad \ldots\ldots(2.2)$$

Effects associated with more complex sources are described in Part III of this series (Cowley and Moodie 1957 b).

For the purposes of the present discussion the source function is a delta function located on the axis. Making the necessary substitutions the equations reduce to

$$u=\exp\left\{-ik\frac{x^2}{2R}\right\}\left[F\left(-k\frac{x}{R}\right)\star\exp\left\{ik\frac{R_q}{2R(R+R_q)}x^2\right\}\right],$$

$$u=\exp\left\{-ik\frac{x^2}{2R}\right\}\exp\left\{ik\frac{R_q}{2R(R+R_q)}x^2\right\}$$
$$\times\left[\rho\left(\frac{R_q}{R+R_q}x\right)\star\exp\left\{-ik\frac{R_q}{2R(R+R_q)}x^2\right\}\right].$$

Since only the intensity distribution is observable, the terms of modulus unity outside the square brackets may be neglected for the present, so that the equations to be considered are

$$u=F\left(-k\frac{x}{R}\right)\star\exp\left\{ik\frac{R_q}{2R(R+R_q)}x^2\right\}, \qquad \ldots\ldots(2.3)$$

$$u=\rho\left(\frac{R_q}{R+R_q}x\right)\star\exp\left\{-ik\frac{R_q}{2R(R+R_q)}x^2\right\}. \qquad \ldots\ldots(2.4)$$

§ 3. PYSICAL INTERPRETATION OF THE EQUATION

For the general system involving Fraunhofer diffraction the wave function is expressed as a function or functions of reciprocal space. For the general system involving Fresnel diffraction, however, it has been shown (Cowley and Moodie 1957 c) that the wave function may be expressed as a combination of functions of direct space, reciprocal space, or direct and reciprocal space. The particularly simple case described by equations (2.3) and (2.4) offers a clear example of this. Thus equation (2.3) expresses the wave function as a convolution of two functions of reciprocal space, and equation (2.4) expresses the same function as a convolution of two functions of direct space.

Specifically, equation (2.3) expresses the wave function as the Fourier transform of the object convoluted with the Fourier transform of a Fresnel wave function.† If the object is periodic its Fourier transform consists of a series of equally spaced Dirac delta functions and the convolution reduces to a summation. In fact the Fourier transform of a Fresnel wave function is laid down at each delta function with the weight and sign of that function, and the whole summed.

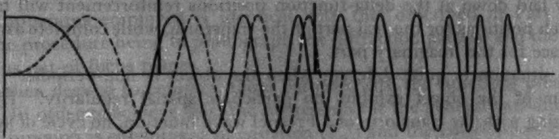


Figure 1 (*a*). Diagrammatic representation of an out-of-focus condition. The full line represents $\cos x^2$ and the broken line $\sin x^2$. Vertical lines represent the weighted delta functions which comprise the transform of a periodic object.

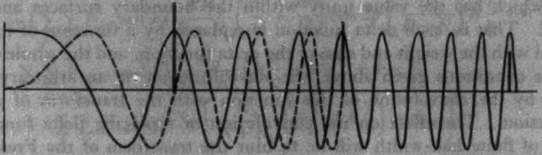


Figure 1 (*b*). Diagrammatic representation of the focal condition for the second Fourier image.

The process is represented diagrammatically in figures 1(*a*), (*b*), where the functions $\sin x^2$ and $\cos x^2$ are sketched along with a lattice of delta functions. In order to avoid confusion the continuous functions are not laid down at each delta function. Figure 1(*a*) represents the position on an arbitrary plane of observation. No obvious relationship exists between the positions of the delta functions and the maxima and minima of the real and imaginary parts of the Fourier transform of the Fresnel wave function. Hence if Fresnel wave functions are laid down at each of the delta function positions they will tend to cancel out and an ill-defined pattern, typical of out-of-focus setting, will result. Movement of the plane of observation will shift the positions of the maxima and minima relative to the positions of the delta functions and lead to the possibility of more distinctive reinforcement. When the periodic object has few imperfections and possesses considerable fine detail an extensive array of delta functions with considerable weight in the high orders will result so that the probability of distinctive reinforcement at an arbitrary setting will be increased. In fact it is observed that with gratings of this type a large number of sharp and frequently complicated out-of-focus patterns are generated.

---

† The term 'Fresnel wave function' is used to describe the wave function due to a point source, since convolution with this function leads to the Fresnel fringe distributions seen, for example, in the out-of-focus images from small sources.

As the plane of observation continues to be moved, more obvious relationships between the positions of the maxima and minima and those of the delta functions arise. An extreme example of this is illustrated in figure 1(*b*) where certain maxima of $\cos x^2$ and the corresponding zeros of $\sin x^2$ are seen to coincide with the positions of delta functions. Specifically the sequence of coincidences of the maxima with delta function positions is that of the squares of the integers. This is in effect the focal condition for the pattern $I(1)$, the pattern on the plane of the second Fourier image, and figure 1(*b*) shows in diagrammatic form how this pattern is formed. It can be seen that if transforms of Fresnel wave functions are now laid down at the delta-function positions reinforcement will take place at all such positions for the real part of the transform, while complete cancellation takes place for the imaginary part.

The situation is quite different when the object is non-periodic. Now the transform of the object exhibits no particular spatial regularity. Hence, on convoluting with the transform of a Fresnel wave function, no cycle of systematic reinforcement is possible, so that, on moving the plane of observation away from the object, the pattern merely becomes increasingly blurred. The finite periodic object affords an interesting special case. The transform of this object is given by the convolution of the lattice of delta functions with the transform of the function which has the value unity within the boundary surfaces and is zero elsewhere. That is, each delta function is replaced by a function of finite half-width, and with the weight and sign of the delta function, and the whole summed. Under the conditions given above, the wave function on an arbitrary plane is described by the convolution of this function with the transform of a Fresnel wave function. The effect on the convolution of replacing delta functions by functions of finite half-width will be to blur the transform of the Fresnel wave function, particularly at high values of the argument. Hence the contributions to the image of the high orders of diffraction will be virtually destroyed, with a corresponding reduction in resolution. It is thus apparent that for arbitrary bounds in two dimensions it will no longer be possible to satisfy the conditions for focus exactly, that resolution will be reduced, and that this will be most pronounced at the bounds of the pattern. Further it is clear, as might be expected on general grounds, that resolution deteriorates with decreasing size of object, and the importance of this may be rapidly estimated by an order of magnitude calculation. Other cases of interest, for instance the periodic object with prescribed faulting, may be treated quantitatively by a similar procedure.

Equation (2.4) describes the wave function as the convolution of the object with a Fresnel wave function. This description corresponds directly to what is actually observed experimentally with a microscope. An account of this has already been given in Part I of this series (Cowley and Moodie 1956 a) and will not be repeated. The form of the equation, however, emphasizes several important features of the phenomenon. Thus the argument of the object function shows directly that a magnification corresponding to that for a shadow image will be observed. Further, it is clear that if the object exhibits periodicity a regular cycle of reinforcement will occur in direct space, while if no periodicity is present only dissipation of the Fresnel wave functions is possible.

At first sight the imaging processes described in this series appear to be divorced from those usually encountered in optics, and the methods employed in describing them appear to be correspondingly specialized. This is not, however, the case. A description of imaging by standard light lens systems, including phase contrast systems, may be obtained by appropriate substitution in, and manipulation of, equation (2.1) and its companions (Cowley and Moodie 1957 c). What is common to all of these systems is the presence of a function, self-reciprocal under a Fourier transformation, in an equation of the form (2.1). In the present instance the self-reciprocal function is $\exp\{-iC\xi^2\}$, a Fresnel wave function. This matter will be discussed in some detail in a later publication.

§ 4. EXPERIMENTAL

The apparatus employed is identical with that described in Part I of the series, the procedure being to photograph the diffraction pattern and the patterns on various out-of-focus planes.

It is then desired to compare these patterns with the functions predicted by theory. In order to achieve this, the expression for the general two-dimensional periodic function is substituted in equation (2.3) and the convolution evaluated. There results

$$\begin{aligned}\rho(x,y) = {} & \sum_h \sum_k F_{hk} \cos \pi\lambda \frac{RR_q}{R+R_q}\left(\frac{h^2}{a^2}+\frac{k^2}{b^2}\right) \cos 2\pi \left[\frac{R_q}{R+R_q}\left(h\frac{x}{a}+k\frac{y}{b}\right)+\phi_{hk}\right] \\ & + i \sum_h \sum_k F_{hk} \sin \pi\lambda \frac{RR_q}{R+R_q}\left(\frac{h^2}{a^2}+\frac{k^2}{b^2}\right) \cos 2\pi \left[\frac{R_q}{R+R_q}\left(h\frac{x}{a}+k\frac{y}{b}\right)+\phi_{hk}\right] \\ = {} & \sum_h \sum_k F_{hk}' \cos 2\pi \left[\frac{R_q}{R+R_q}\left(h\frac{x}{a}+k\frac{y}{b}\right)+\phi_{hk}\right] \\ & + i \sum_h \sum_k F_{hk}'' \cos 2\pi \left[\frac{R_q}{R+R_q}\left(h\frac{x}{a}+k\frac{y}{b}\right)+\phi_{hk}\right] \qquad \ldots\ldots(4.1)\end{aligned}$$

where

$$F_{hk}' = F_{hk} \cos \pi\lambda \frac{RR_q}{R+R_q}\left(\frac{h^2}{a^2}+\frac{k^2}{b^2}\right); \quad F_{hk}'' = F_{hk} \sin \pi\lambda \frac{RR_q}{R+R_q}\left(\frac{h^2}{a^2}+\frac{k^2}{b^2}\right).$$

Since the grating employed had a square unit cell, $a=b$. It is further convenient to express F_{hk}' and F_{hk}'' in terms of μ, the fractional order number, i.e.

$$\mu = \frac{1}{2}\frac{\lambda}{a^2}\frac{RR_q}{R+R_q}.$$

Then

$$F_{hk}' = F_{hk} \cos 2\pi\mu(h^2+k^2); \quad F_{hk}'' = F_{hk} \sin 2\pi\mu(h^2+k^2).$$

The patterns $I(0)$, $I(\frac{1}{8})$ and $I(\frac{1}{4})$ were then calculated in the following way. Intensities of the various diffraction orders other than that of zero order were estimated by eye from the diffraction pattern. This process could have been made accurate only at the expense of considerable complication. Since the precise details of the pattern are of no great inherent interest it was considered that an estimate, which should be sufficient to test the ability of the theory to describe the outstanding features of the out-of-focus patterns, would be adequate.

A knowledge of the form of the periodic object, a magnified image of which is shown in figure 2(*a*) (Plate I), allowed the phases of the structure amplitudes to be calculated. The double Fourier summation for $I(0)$ could then be computed on a relative basis by means of Beevers–Lipson strips. The result of this computation is shown in figure 3 and it can be seen that a reasonable representation of the gauze is obtained despite the marked departures from periodicity remarked on in Part I.

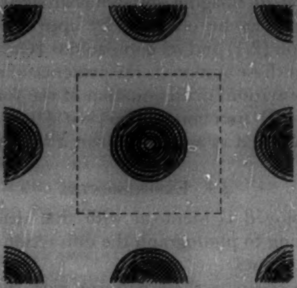


Figure 3. Intensity distribution in an image plane calculated by summation of the Fourier series.

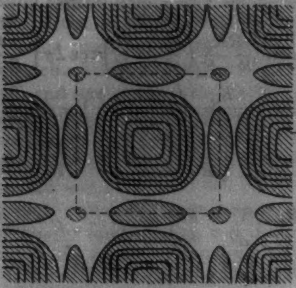


Figure 4. Calculated intensity distribution for *I* (0·125). The broken lines are at half the interval of the full lines.

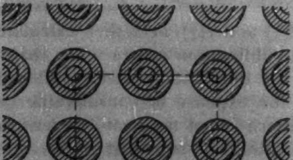


Figure 5. Calculated intensity distribution for *I* (0·25).

Since it is evident that $F_{00}' = F_{00}$, while $F_{00}'' = 0$ it is necessary to put further calculations on an absolute basis by providing an estimate of F_{00}. This is achieved by noting the value which has to be added to the summation for $I(0)$ to make the lowest region non-negative.

The summations for $I(\frac{1}{8})$ and $I(\frac{1}{4})$ were then carried out, the results being shown in figures 4, 5.

In figures 3, 4 and 5 the intensity distribution is mapped, that is, the real and imaginary parts of the summations are squared after the addition of the appropriate constant to the real part as described above, and the two added.

Values of F_{hk}, F_{hk}' and F_{hk}'' appropriate to the various summations are given in the table.

Structure Factors for Various Values of the Fractional Order Number

	$\mu=0$	$\mu=\frac{1}{8}$		$\mu=\frac{1}{4}$	
hk	F_{hk}	F_{hk}'	F_{hk}''	F_{hk}'	F_{hk}''
00	168	168	0	168	0
10	−100	−70	−70	0	−100
20	30	−30	0	30	0
11	60	0	60	−60	0
21	−20	14	14	0	−20
22	10	10	0	10	0

A series of experimentally observed out-of-focus patterns is shown in figures 2(*a*)–(*h*) (Plates I and II). The order numbers of these patterns are, in sequence: 0, 0·075, 0·142, 0·204, 0·258, 0·310, 0·363, 0·413 and 0·538. The symmetry of this cycle between the zero-order and the first-order Fourier image is apparent and in accord with the theory presented.

There are no patterns exactly at order numbers 0·125 and 0·25, but comparison of figures 4 and 5 with patterns of adjacent order number shows that, despite the crudeness of the intensity estimates, the theoretical and experimental results agree in reasonable detail.

Again equation (4.1) predicts that while the symmetry in an out-of-focus pattern may be higher than that of the object, it can never be lower, and figure 2 (Plates) shows that this condition is satisfied in practice.

§ 5. CONCLUSION

Within the limits of experimental accuracy the theory presented in this series appears to give an adequate quantitative description, at any point in the field, of the interaction of a spherical wave with a two-dimensional periodic lattice.

The deficiencies of Kirchhoff's equation and related approximations to the solution of wave equations have been examined by many authors with agreement on the failure of the approximations in certain experimental situations.

However, no quantitative estimate of the range of validity of the screen approximation is available, although doubts have been expressed as to its applicability when interactions at distances of the order of wavelengths from boundaries have to be considered. Equation (2.1) is correct only to the Fresnel approximation which is much weaker than the Kirchhoff and related approximations, and it has been applied to systems involving multiple interactions from apertures separated in some cases by distances of only a few wavelengths.

Nevertheless, over the limited angular range which has been tested, no deviation from predicted behaviour has been observed.

We therefore conclude that some confidence may be felt in applying equations of the type (2.1) to a wider class of optical phenomena than that considered here.

REFERENCES

COWLEY, J. M., and MOODIE, A. F., 1957 a, *Proc. Phys. Soc.* B, 70, 486; 1957 b, *Ibid.*, 70, 505.

Fourier Images : III—Finite Sources

By J. M. COWLEY and A. F. MOODIE
Division of Industrial Chemistry, Commonwealth Scientific and Industrial Research Organization, Melbourne, Australia

Communicated by A. L. G. Rees; MS. received 4th April 1956

Abstract. The theoretical and experimental investigation of Fourier images, described in Parts I and II, is extended to the case of finite sources. The Fourier images produced by a pair of point sources, a rectangular source and a Gaussian source are considered in detail. The dependence of focus on source dimensions for the first two of these cases is confirmed experimentally with light optics.

From the results obtained for these idealized sources, the contrast and resolution of Fourier images from experimentally realizable sources is deduced.

The possibility is considered of an improvement in resolution in the electron microscopy of thin crystals when the first stages of magnification are performed by forming magnified Fourier images. It is shown that the resolution that can be attained depends on the sharpness with which the boundaries of the source are defined. In particular, if the effective source is formed by using electron lenses to demagnify a larger source, the resolution may be improved by a factor of $2^{-1/2}$.

---

§ 1. Introduction

In the previous two papers of this series (Cowley and Moodie 1957 a, b, hereafter referred to as Part I and II) an investigation has been made of the formation of Fourier images when a periodic object is illuminated from a small source of radiation. For the idealized case of a point source and an infinite two-dimensional periodic object it was shown that, within the limits of validity of a parabolic approximation to the spherical wavefronts, a number of perfectly resolved, in-focus, magnified and demagnified images of the object are produced. The principal theoretical conclusions as to the positions and magnifications of the in-focus images and the form of the out-of-focus images have been confirmed experimentally by the use of visible light and photographic gratings with periodicities of the order of $20\,\mu$.

The formation of Fourier images with visible light would appear to be of academic interest only. Objects with periodicities in the micron range form a very limited class. Moreover, the resolving power of optical microscopes approaches the limit set by the wavelength of the light, so that the use of magnified Fourier images would appear to offer no advantage except, perhaps, as a means of distinguishing a periodic component in a principally non-periodic object. In electron microscopy, on the other hand, the resolution is limited by the aberrations of the objective lens. In the electron microscopy of periodic objects, such as crystal lattices, an important practical advantage would result if the first stages of magnification could be made by forming magnified Fourier images in which the loss of resolution is less than that in an objective lens.

If a point source of electrons could be used, there would be no loss of resolution in the Fourier images. The point source approximation applies, however, only when the source diameter is very much less than the periodicity of the object, i.e. in this case, less than 1 Å. This is not readily realizable in practice. It is necessary, therefore, to determine what loss of resolution, if any, results from the use of of finite sources, such as can be produced experimentally. In this paper we approach this problem by obtaining approximate expressions for the Fourier images produced with various idealized forms of finite source. Fairly definite conclusions can be drawn from these results regarding the effects of realizable sources. As before, some of the principal theoretical results have been confirmed by optical experiments.

For convenience we assume that all sources considered are coherent and that each element of the source emits radiation uniformly in all directions. The periodic object will be considered to have infinite extent. The analysis will be limited to one-dimensional objects. The extension to two dimensions follows readily.

As in Part II of this series, we base our analysis on formulae derived from the unified theory of wave optics described elsewhere (Cowley and Moodie, to be published), and in particular on the expression for the amplitude in the plane of observation

$$u=K\exp\left(\frac{-ikx^2}{2R}\right)\left[F\left(-\frac{kx}{R}\right)\star\exp\left\{\frac{ikR_qx^2}{2R(R+R_q)}\right\}\star\ q\left(\frac{R_qx}{R}\right)\exp\left\{\frac{-ikR_qx^2}{2R^2}\right\}\right] \qquad \ldots\ldots(1.1)$$

where K is a constant, $q(x_q)$ is the source function, and $F(x)$ is the Fourier transform of the object transmission function $\rho(\xi)$. The notation is that of Part I.

If $\rho(\xi)$ is expressed as a Fourier series,

$$\rho(\xi)=\sum_h F_h\cos 2\pi\frac{h}{a}\xi$$

the first convolution in (1.1) can be evaluated, giving

$$\begin{aligned}u=K\exp&\left(-\frac{ikx^2}{2R}\right)\\&\times\left[\left\{\exp\left(\frac{ikR_qx^2}{2R(R+R_q)}\right)\sum_h F_h\exp\left(\frac{i\pi RR_q\lambda h^2}{(R+R_q)a^2}\right)\cos\frac{2\pi R_qhx}{(R+R_q)a}\right\}\star\left\{q\left(-\frac{R_qx}{R}\right)\right.\right.\\&\left.\left.\times\exp\left(-\frac{ikR_qx^2}{2R^2}\right)\right\}\right].\end{aligned} \qquad \ldots\ldots(1.2)$$

Thus the amplitude in the plane of observation can be expressed as the convolution of the amplitude due to a point source with the product of the source function and an exponential term.

§ 2. Two Point Sources

This case is a useful intermediate between the point-source case and that of various finite sources, introducing, as it does, the dependence of focusing properties on source dimensions. The separation of the two point sources is taken as 2η, so that the source function is

$$q(x_q)=\delta_{x_q=\eta}+\delta_{x_q=-\eta}.$$

Performing the convolution in (1.2) then gives

$$u_2 = K\exp\left(-\frac{ik(x^2+\eta^2)}{2(R+R_q)}\right)\sum_h F_h \exp\left(\frac{i\pi RR_q\lambda h^2}{(R+R_q)a^2}\right)$$
$$\times\left[\cos\frac{2\pi R_q hx}{(R+R_q)a}\cos\frac{2\pi Rh\eta}{(R+R_q)a}\cos\frac{2\pi}{\lambda}\frac{x\eta}{(R+R_q)}\right.$$
$$\left.+i\sin\frac{2\pi R_q hx}{(R+R_q)a}\sin\frac{2\pi Rh\eta}{(R+R_q)a}\sin\frac{2\pi}{\lambda}\frac{x\eta}{(R+R_q)}\right].$$

Provided that the separation of the point sources is not too large ($\eta \ll R\lambda/a$) the term $\cos 2\pi x\eta/(R+R_q)\lambda$ does not differ appreciably from unity for values of x corresponding to many periods of the magnified image. Hence for the paraxial image we may write

$$u_2 = K\exp\left(-\frac{ik(x^2+\eta^2)}{2(R+R_q)}\right)\sum_h F_h \exp\left(\frac{i\pi RR_q\lambda h^2}{(R+R_q)a^2}\right)\cos\frac{2\pi Rh\eta}{(R+R_q)a}\cos\frac{2\pi R_q hx}{(R+R_q)a}. \qquad \ldots\ldots(2.1)$$

In order that the image should be an exact magnified reproduction of the object, two conditions must be satisfied, namely

$$\frac{RR_q}{R+R_q} = \frac{2\nu a^2}{\lambda} \quad \text{and} \quad \eta = \frac{m(R+R_q)a}{R}, \qquad \ldots\ldots(2.2)$$

where m and ν are integers or half-integers. The first condition defines the set of focal planes, as for the case of a single point source. The second condition restricts the separation of the point sources to multiples of a basic distance. It may be noted that this is the condition that the images of the grating produced by the two point sources considered separately should coincide.

These conclusions were tested experimentally with visible light, using the apparatus described in Part I. Two small sources of light of variable separation were obtained by using the microscope condenser lens to image the regions of intersection of a single illuminated slit and a double slit in the form of a narrow V which could be translated at right angles to the single slit. Figures 1 (*a*)–(*e*) (Plate I) illustrate the changes occurring in the first real Fourier image ($\mu = +\frac{1}{2}$) of the grating shown in figure 3 of Part I, as the separation of the pair of sources is increased from 17 to 143 μ in approximately equal steps. The individual sources were approximately rectangular and 20 μ by 10 μ, the smaller dimension being in the direction of their separation. The Fourier image is seen to be in focus for source separations of zero and about 150 μ. The difference in separation for two in-focus positions calculated from equation (2.2) is 168 μ. The out-of-focus images are similar to, and may be calculated in the same way as, the out-of-focus images (2μ not integral) from a single point source described in Part II. The results obtained for two point sources also apply to arrays of equally spaced point sources. These, and similar matters, will be considered in a future publication.

§ 3. RECTANGULAR SOURCE

The source function for the one-dimensional equivalent of a rectangular aperture may be written

$$q(x_q) = 1 \text{ for } |x_q| \leqslant \eta_0, \qquad \text{and zero elsewhere.}$$

The effect of this source is obtained by integrating the amplitude function (2.1) over values of η from 0 to η_0 thus

$$u_3 = K\exp\left(\frac{-ikx^2}{2R}\right)\sum_h F_h \exp\left(\frac{i\pi RR_q\lambda h^2}{(R+R_q)a^2}\right)\cos\frac{2\pi R_q hx}{(R+R_q)a}\int_0^{\eta_0}\exp\left(\frac{-ik\eta^2}{R+R_q}\right)$$

$$\times\cos\frac{2\pi Rh\eta}{(R+R_q)a}\,d\eta. \qquad \ldots\ldots(3.1)$$

The integral over η can be evaluated explicitly in terms of Lommel's functions. However, equation (2.1) already contains the approximation that η is small compared with R. With a similar restriction, namely $\eta^2 \ll (R+R_q)\lambda$, the exponential term in the integral can be taken as approximately unity over the range of integration. The amplitude function then becomes identical with that due to a point source except that the Fourier coefficients, or structure amplitudes, F_h, are multiplied by

$$\left(1\Big/\frac{2\pi Rh}{(R+R_q)a}\right)\sin\frac{2\pi Rh\eta_0}{(R+R_q)a}. \qquad \ldots\ldots(3.2)$$

Because the maxima and minima of the function $(\sin y)/y$ do not occur at multiples of any particular y value, there is no value of η_0 for which all F_h are multiplied by unity or by $(-1)^h$. Hence it is not possible to get an exact image of the grating except in the limiting case of $\eta_0 \ll a$. However, if η_0 is chosen such that

$$\eta_0 = \frac{(R+R_q)}{R}\,a\,(m+\epsilon) \qquad \ldots\ldots(3.3)$$

where m is an integer, and ϵ is sufficiently small, so that $hm\epsilon \ll 1$ for all h for which F_h is not negligible, then the expression (3.2) becomes $\eta_0\epsilon(/1+\epsilon)$ for $h \neq 0$, and η_0 for $h=0$. Hence, neglecting terms of modulus unity and constants, the amplitude function becomes

$$u_3 = \eta_0\left[F_0(1-\epsilon)+\epsilon\sum_h F_h\exp\left(\frac{i\pi RR_q\lambda h^2}{(R+R_q)a^2}\right)\cos\frac{2\pi R_q hx}{(R+R_q)a}\right]. \qquad \ldots\ldots(3.4)$$

The image is therefore identical with that given by a very small source of width $a(R+R_q)\epsilon/R$, except for the addition of a large constant term. The image contrast is thus reduced by a factor of approximately $1/\epsilon$ but the resolution is unimpaired on the focal planes.

For the source sizes given by $\eta_0 = (R+R_q)\,a\,(m+\frac{1}{2}+\epsilon)/R$ and ϵ very small, as before, the F_h are multiplied by $(-1)^h\epsilon$. The image is then identical with (3.4) except for a translation of a half of one period.

The effect of increasing ϵ is illustrated in figure 2. For small ϵ it can be seen that the values of $(\sin y)/y$ for $y = n\pi(1+\epsilon)$ or $2n\pi(1+\epsilon)$ are almost constant for n not too large. For larger ϵ the values of $[\sin 2n\pi(1+\epsilon)]/2n\pi(1+\epsilon)$ are initially larger but fall off rapidly for higher values. As ϵ increases, therefore, the initial effect is that the F_h for h large are multiplied by factors less than ϵ. The decrease becomes important at about $hm\epsilon = \frac{1}{4}$. Hence, while the contrast of the image increases, the resolution decreases.

For still larger ϵ, the multipliers of the F_h alternate in sign, and, on the average, decrease in proportion to $1/h$. The image is therefore distorted and the resolution is poor.

In order to obtain a high-resolution Fourier image in the two-dimensional case of a rectangular aperture, a relation of the form (3.3) must be satisfied by both aperture dimensions with respect to the appropriate unit-cell parameters of the grating and the values of R_q and R must be chosen to give focusing conditions in both directions.

Optical experiments were made using a rectangular source produced by imaging the region of intersection of an illuminated slit and a narrow, triangular aperture. One dimension of the source could be varied over a wide range by translating the triangular aperture relative to the slit. A Fourier image was brought into focus with the source as small as possible. As one of the source dimensions was increased, the contrast of the corresponding periodic component of the grating could be observed to vary from zero, corresponding to $\epsilon = 0$, to a maximum at $\epsilon = \frac{1}{2}$. The predictions as to the variation in resolution could be confirmed qualitatively.

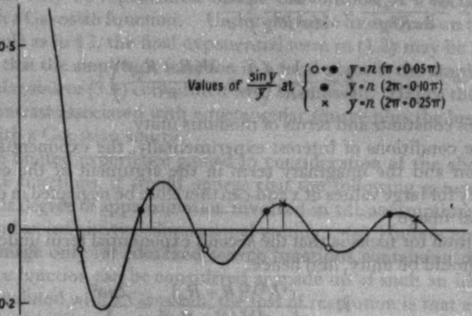


Figure 2. Plot of $(\sin y)/y$, showing the variation of the values of the function at $y = n\pi(1 \pm \epsilon)$ or $2n\pi(1 \pm \epsilon)$, with n and ϵ.

The series of photographs, figures 3(*a*)–(*f*) (Plate II) were taken of the $\nu = 2\frac{1}{2}$ Fourier image, of the 14·67 μ grating used for figure 7 of Part I and figure 2 of Part II. The exposures were made for values of the variable dimension of the rectangular source corresponding to maximum contrast in the corresponding fringes. The other source dimension was fixed at a value such that the corresponding periodic component of the image had fairly low contrast. For each exposure of the Fourier image the size of the source was recorded by focusing the microscope on the plane of the source and re-exposing to add an image of the diffraction pattern to the plate.

For this particular grating, the diffraction pattern (figure 7(*a*), Part I), is very limited and is dominated by strong first-order reflections. The modulating factor (3.2) therefore has little effect on the image for the maximum contrast ($\epsilon = \frac{1}{2}$) positions. In the table, the source dimensions measured from the plates of figure 3 are compared with those calculated from the relation (3.3). There is satisfactory agreement within the limits of experimental accuracy.

Comparison of Observed and Calculated Dimensions of a Rectangular Slit for Maximum Contrast (in μ)

m	$(2\eta_s)_{\mathrm{obs}}$	$(2\eta_s)_{\mathrm{calc}}$
0	<13	0
$\frac{1}{2}$	40	36
1	64	59
$1\frac{1}{2}$	92	83
2	115	106
$2\frac{1}{2}$	136	130

§ 4. GAUSSIAN SOURCE

The source function is taken as $q(x_q)=\exp(-cx_q^2)$, a Gaussian function with width, at a value $1/e$ times the maximum, of $\sigma=2/c$. Equation (1.2) then reduces to

$$u=K\exp\left(\frac{-x^2}{16c^2R^2\chi}\right)\sum_h F_h\exp\left(\frac{-\lambda^2h^2}{16c^2R^2\chi a^2}\right)\exp\left\{i\left(\pi\lambda R-\frac{(R+R_q)\lambda^3}{16\pi R^2\chi}\right)\frac{h^2}{a^2}\right\}$$
$$\times\cos\left\{2\pi-\frac{(R+R_q)\lambda^2}{8\pi R^3\chi}-\frac{i\lambda}{16c^2R^3\chi}\right\}\frac{hx}{a} \qquad \ldots\ldots(4.1)$$

where

$$\chi=\frac{1}{16R^4}\left(\frac{1}{c^4}+\frac{\lambda^2(R+R_q)^2}{\pi^2}\right)$$

and K includes constants and terms of modulus unity.

Under the conditions of interest experimentally, the exponential term before the summation and the imaginary term in the argument of the cosine become important only for large values of x and can therefore be neglected in considerations of the paraxial image.

The condition for focus is that the second exponential term under the summation in (4.1) should be unity, and hence

$$R-\frac{c^4R^2\lambda^2(R+R_q)}{\pi^2+c^4\lambda^2(R+R_q)^2}=\frac{2na^2}{\lambda}$$

where n is an integer. For c very large, i.e. the source width σ very small, this becomes $RR_q/(R+R_q)=2na^2/\lambda$, as for a point source. For very large σ, so that very nearly parallel illumination falls on the object, the condition is $R=2na^2/\lambda$, as for a point source at infinity.

The image in a plane of focus then becomes

$$u=K\sum_h F_h\exp\left(\frac{-\lambda^2h^2}{16c^2R^2\chi a^2}\right)\cos\left\{2\pi-\frac{(R+R_q)\lambda^2}{8\pi R^3\chi}\right\}\frac{hx}{a} \qquad \ldots\ldots(4.2)$$

This represents an image of the object with magnification

$$M=\frac{\pi^2+\lambda^2c^4(R+R_q)^2}{\pi^2+\lambda^2c^4R_q(R+R_q)}=\frac{\pi^2\sigma^4+16\lambda^2(R+R_q)^2}{\pi^2\sigma^4+16\lambda^2R_q(R+R_q)}.$$

The exponential term is of the nature of an 'artificial temperature factor' decreasing the resolution. The magnified image is thus convoluted with the transform of this factor, which is another Gaussian. The width of this Gaussian is a measure of the least resolvable distance in the image. Dividing this width by the magnification gives the least resolvable distance referred to the scale of the object as

$$\Delta=\frac{4R\sigma\lambda}{[\pi^2\sigma^4+16\lambda^2(R+R_q)^2]^{1/2}}.$$

Hence, for σ small, the magnification is very close to that for a point source, viz. $(R+R_q)/R_q$, and the least resolvable distance Δ is very nearly equal to $R\sigma/(R+R_q)$. As σ increases, Δ passes through a maximum for $4\pi\sigma^2=(R+R_q)\lambda$ and then begins to decrease while, in approximately the same range of σ, M falls from near the point-source value to near unity. For very large σ, the least resolvable distance Δ decreases in proportion to $1/\sigma$ and may become very small, but the magnification M does not differ appreciably from unity.

The difficulty of producing a light source with a Gaussian amplitude distribution has precluded experimental tests of these conclusions.

§ 5. Other Sources

It is not difficult to calculate the effect of sources which can be considered as combinations of the ideal sources treated above. Perhaps the most important of these, from a practical viewpoint, is the rectangular source with diffused edges, as in the image of an aperture produced by an imperfect lens, or with uneven edges. Such a source may be represented by the convolution of a rectangular source function with a Gaussian function. Under the same limitations on the dimensions of the rectangle as in § 3, the final exponential term in (1.2) may be neglected. It then follows that the amplitude function in a focal plane is the amplitude function of a rectangular source (3.4) convoluted with a Gaussian. Hence the image suffers the loss of contrast associated with a rectangular source plus the loss of resolution associated with a Gaussian source.

From the limited experience gained in consideration of the above and other simple source functions, it would appear that the following general conclusions are valid to the degree of approximation involved in the assumptions of a parabolic wave front and an infinite two-dimensional grating: (*a*) There is no loss of resolution in the image only for idealized source functions containing discontinuities. (*b*) If a source function can be considered as made up of such an idealized source function convoluted with a Gaussian, the loss of resolution is that associated with the Gaussian. (*c*) Sources of infinite width, with no discontinuities, give images of unit magnification. (*d*) If the source contains a periodic component, this component will appear in the Fourier images. This becomes apparent when equation (1.1) is expressed in a form symmetrical with respect to the source and the object (Cowley and Moodie, to be published).

So far we have considered only a real source function. This restriction is by no means necessary. The basic equations (1.1) and (1.2) are equally applicable for complex or imaginary source functions. Similarly the transmission function of the object may be either real or complex.

§ 6. Factors Limiting Image Resolution

It has been assumed in the above that all sources are coherent. Partially coherent sources could be considered if, following Hopkins (1951), a 'phase-coherence' factor were introduced. However it can be shown that, for a self-luminous or other incoherent source, the in-focus images are at least qualitatively the same as for a completely coherent source in the case of the real sources considered above, provided that the source dimensions are not too large. In particular, for an incoherent Gaussian source, $\Delta=R\sigma/(R+R_q)$ for all σ, i.e., the least resolvable distance increases in proportion to the source diameter, whereas for a coherent source $\Delta = R\sigma/(R+R_q)$ for small σ but falls appreciably below this value for σ values of the order of $\{(R+R_q)\lambda/4\pi\}^{1/2}$. For coherent and incoherent rectangular sources for which $\eta_0^2 \ll (R+R_q)\lambda$, the in-focus images differ only in contrast. This becomes obvious when it is considered that the intensity distribution may be expressed in terms of a Fourier series, and the image due to a rectangular incoherent source is obtained by convolution of this series with the appropriately magnified or demagnified source function.

The finite frequency range of practical 'monochromatic' sources of light or electrons can readily be shown to make negligible difference to the form or resolution of Fourier images.

In the derivation of the basic equations such as (1.1), a paraboloidal approximation to spherical wave fronts has been made. This approximation is valid for values of x_q, ξ and x much less than R_q and R. When it is applied in the case of an infinite two-dimensional grating, the image amplitude functions are in error to the extent that they contain no wavelength limitation to the resolution. The use of the approximation in this case can be justified only if the structure amplitudes F_h of the grating fall off sufficiently rapidly with increasing h, i.e. if the wavelength of the radiation is much smaller than any detail in the object.

In practice the grating will be of finite extent, and the aperture of the illumination from the source may be limited so that not all of the grating will be illuminated. The loss of resolution in the image due to the limitation of the dimensions of the grating will be identical with that due to a similarly limited lens, of appropriate focal length, replacing the grating and imaging a grating in the plane of the source. Since, however, the magnification of the lens would be R/R_q whereas that of the grating would be $(R+R_q)/R_q$, the least resolvable distance Δ, referred to the grating itself, would be less by a factor of $R/(R+R_q)$ for the Fourier image produced by the grating than for the image produced by the lens.

§ 7. APPLICATION TO THE ELECTRON MICROSCOPY OF CRYSTALS

The possibility that in the electron microscopy of thin crystals the resolution could be improved by obtaining the first stages of magnification by Fourier image formation may now be considered in detail. For this purpose a small source of electrons must be provided at a distance from the specimen not too many times greater than $2a^2/\lambda$. Thus for electrons of about 50 kv and periodicities of the order of 10 Å, R_q will be a few microns. The diameter of the electron source may be much greater than the least resolvable distance required in the image. It was shown for a rectangular source that, if the focusing conditions for η_0 and R were satisfied, the resolution was not affected by increasing the source size, but the contrast was decreased. If however the undeflected electrons could be prevented from reaching a portion of the image, i.e. if, for this portion of the image, the F_0 term in the expression (3.4) could be reduced to zero, no loss of contrast would result.

Suitable sources of electrons could be produced by (*a*) using small electron-emitting particles, (*b*) illuminating a small aperture with an intense electron beam, (*c*) using electrons elastically scattered from a small, thin particle or from the edges of a thick particle or an aperture, (*d*) using an electron-optical lens to form a demagnified image of a small source or aperture.

If the Fourier image has a magnification M and the least resolvable distance referred to the crystal lattice is Δ_F, and the Fourier image is then magnified further by a microscope system which can resolve a distance of Δ_{m}, the final image will have a least resolvable distance, referred to the crystal lattice, of $\Delta = \{\Delta_{\mathrm{f}}^2 + (\Delta_{\mathrm{m}}/M)^2\}^{1/2}$. If the conditions for focus are satisfied, $\Delta_{\mathrm{f}} = R\sigma/(R+R_q)$, where σ is a measure of the unevenness of the edges for sources of types (*a*), (*b*) and (*c*) above, and a measure of the loss of resolution due to the aberrations of the electron lens used in (*d*).

For the case (*d*), the best resolution is obtained when the aberrations of the lenses used to demagnify the source and to magnify the image are the same and as small as possible. Then $\sigma = \Delta_{\mathrm{m}}$ and $\Delta = \Delta_{\mathrm{m}}(R^2 + R_q^2)^{1/2}/(R+R_q)$ and Δ has a minimum value of $\Delta_{\mathrm{m}}/2^{1/2}$ when $R = R_q$ and $M = 2$. For the other types of source, the optimum magnification is $(\Delta_{\mathrm{m}}^2 + \sigma^2)/\sigma^2$, and the corresponding least resolvable distance is $\Delta = \sigma\Delta_{\mathrm{m}}/(\sigma^2 + \Delta_{\mathrm{m}}^2)^{1/2}$. It therefore appears that, in order to obtain any marked improvement in the resolution of electron microscope images of crystal lattices, it is necessary to produce an electron source in one of the ways (*a*), (*b*) or (*c*) above, such that σ is considerably less than the Δ_{m} for the best available electron microscopes.

REFERENCES

COWLEY, J. M., and MOODIE, A. F., 1957 a, *Proc. Phys. Soc.* B, **70**, 486 ; 1957 b, *Ibid.*, **70**, 497.

HOPKINS, H. H., 1951, *Proc. Roy. Soc.* A, **208**, 263.

Paramagnetic Resonance Absorption in Gadolinium Trichloride

By C. A. HUTCHISON, Jr.,† B. R. JUDD and D. F. D. POPE
The Clarendon Laboratory, Oxford

Communicated by B. Bleaney; MS. received 12th November 1956, and in final form 4th February 1957

Abstract. The paramagnetic resonance spectrum in crystals of $GdCl_3$ diluted with $LaCl_3$ and $CeCl_3$ has been examined. The splitting of the ground level $^8S_{7/2}$ of Gd^{3+} in zero magnetic field is small, and has a remarkable temperature dependence. A theory to account for the experimental results is developed, and it is shown that perturbation mechanisms both linear and quadratic in the crystalline field potential are important.

§ 1. Introduction

Paramagnetic resonance absorption in single crystals of $LaCl_3$ containing ions with the electronic configuration f^3 has recently been studied (Hutchison, Llewellyn, Wong and Dorain 1956, Hutchison, Wong and Dorain, to be published, Bleaney, Hutchison, Llewellyn and Pope 1956, Dorain, Hutchison and Wong 1957). In the present paper the paramagnetic resonance of Gd^{3+} (configuration $4f^7$) in $LaCl_3$ is discussed and compared with results previously obtained for Gd^{3+} in other crystal structures (Bleaney, Elliott, Scovil and Trenam 1951, Trenam 1953, Bleaney, Scovil and Trenam 1954).

The Hund ground state of Gd^{3+} is $^8S_{7/2}$ and the theory of the perturbations of the states of this ion by the crystalline field is quite different from that for other ions with f^n configurations. S states are, of course, not split in first order by crystalline fields, and the mechanisms whereby such splittings may occur have been considered by several workers (Van Vleck and Penney 1934, Pryce 1950, Abragam and Pryce 1951, Elliott and Stevens 1953). The results reported in the present paper give a clearer insight into the nature of the mechanisms for the crystalline perturbations in the case of the configuration f^7.

§ 2. Experimental Methods

The method of Anderson and Hutchison (1955) was used to prepare $GdCl_3$ and $LaCl_3$ separately; mixed crystals, containing 0·01 mole per cent $GdCl_3$ were prepared in the manner described by Dorain *et al.* (1957) for UCl_3. Measurements on the paramagnetic resonance absorption were made at approximately 3 cm wavelength, and at room temperature, 90°K and 20°K. $LaCl_3$ has a hexagonal structure, determined by Zachariasen (1948), and the point symmetry at the La^{3+} site is C_{3h}. The peak positions in the resonance absorption pattern with the c axis of the crystal parallel and perpendicular to $\mathbf{H}$ were measured using proton resonance. At intermediate angles, the peak positions were measured by interpolation between proton resonance markers, using photographs of the oscilloscope screen. The same interpolation method was used for a series of measurements made at various angles, keeping the c axis always perpendicular to $\mathbf{H}$.

---

† John Simon Guggenheim Memorial Foundation Fellow 1955–1956.

§ 3. Experimental Results

The spin-Hamiltonian for zero magnetic field must show the same symmetry as the symmetry at a site where a Gd^{3+} ion is located; in this case C_{3h}. Elliott and Stevens (1953) have suggested the spin-Hamiltonian

$$\begin{aligned}\mathscr{H}=\beta\mathbf{H}\,.\,g\,.\,\mathbf{S}&+hcB_2^0\{3S_z^2-S(S+1)\}\\&+hcB_4^0\{35S_z^4-[30S(S+1)-25]S_z^2-6S(S+1)+3S^2(S+1)^2\}\\&+hcB_6^0\{231S_z^6-105S_z^4[3S(S+1)-7]+S_z^2[105S^2(S+1)^2\\&-525S(S+1)+294]-5S^3(S+1)^3+40S^2(S+1)^2-60S(S+1)\}\\&+\tfrac{1}{2}hcB_6^6\{(S_x+iS_y)^6+(S_x-iS_y)^6\}\qquad\ldots\ldots(1)\end{aligned}$$

for a description of the resonance results. The spin angular momentum operators in this Hamiltonian are simply operators with the same transformation properties as the corresponding harmonics Y_n^m necessary for the expansion of the C_{3h} crystalline field potential in the case of f-electrons. The question as to how the B coefficients are related to the crystal field is just the problem of the mechanism of the splitting mentioned above. Following Bleaney *et al.* (1954), we define

$$b_2^0=3B_2^0,\qquad b_4^0=60B_4^0,\qquad b_6^0=1260B_6^0,\qquad b_6^6=1260B_6^6.$$

In terms of these parameters and of $g_{\parallel}$ and $g_{\perp}$, expressions for the energies of the system may be found; and, conversely, from the measured paramagnetic resonance field strengths at constant frequency, the values of the parameters g and b may be determined. This spin-Hamiltonian adequately describes all the experimental data with the values of the parameters given in table 1. A plot of the experimental absorption peak positions against the angle of the c axis with respect to **H** is shown in the figure.

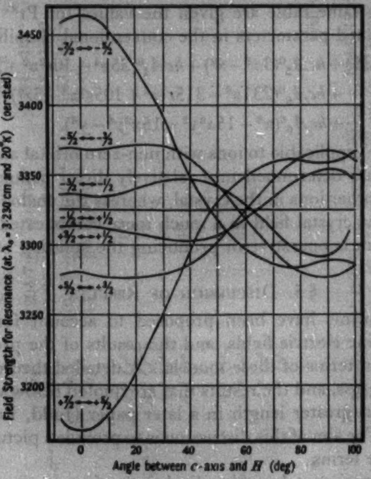


Table 1. Values of the Parameters in the Spin-Hamiltonian

	90°K	Room temperature (~ 290°K)
g	$1{\cdot}991 \pm 0{\cdot}001$	$1{\cdot}991 \pm 0{\cdot}001$
b_2^0	$+(16{\cdot}0 \pm 0{\cdot}02) \times 10^{-4}$ cm^{-1}	$(8{\cdot}36 \pm 0{\cdot}10) \times 10^{-4}$ cm^{-1}
b_4^0	$+(2{\cdot}13 \pm 0{\cdot}05) \times 10^{-4}$ cm^{-1}	$(1{\cdot}68 \pm 0{\cdot}04) \times 10^{-4}$ cm^{-1}
b_6^0	$+(0{\cdot}25 \pm 0{\cdot}05) \times 10^{-4}$ cm^{-1}	$(0{\cdot}64 \pm 0{\cdot}15) \times 10^{-4}$ cm^{-1}
b_6^6	$+(1{\cdot}40 \pm 0{\cdot}3) \times 10^{-4}$ cm^{-1}	

The values of b_2^0, b_4^0 and b_6^0 were calculated from the spacings between corresponding lines on either side of the central line, which eliminates certain asymmetries arising from second-order effects. The value of b_6^6 was obtained in two ways: (*a*) with the c axis always perpendicular to $\mathbf{H}$, the direction of $\mathbf{H}$ was varied and b_6^6 calculated from the variation of the line positions; (*b*) with the c axis perpendicular to $\mathbf{H}$, the value of $b_6^6 \cos\phi - 5b_6^0$ was determined directly from the peak position for an extremum of $\cos\phi$ (see Bleaney *et al.* 1954). Then this value, together with the value of b_6^0 determined in the parallel orientation, was used to obtain b_6^6. The two results were in good agreement.

The signs of the parameters b were determined from the relative intensity measurements in the parallel position at 20°K. The crystals used contained 1·0 mole per cent $\mathrm{CeCl_3}$ in order to avoid saturation. The highest field line was more intense than the lowest field line, and this showed that the value of b_2^0 was a positive number. There was a relatively small change in the values of the b_n^m between 90°K and 20°K. The g values showed no variation with angle or temperature within the errors of the measurements.

§ 4. COMPARISON WITH PREVIOUS RESULTS

A comparison with previous measurements on $\mathrm{Gd^{3+}}$ in other salts is given in table 2. In the same table are given the values for $\mathrm{Pr^{3+}}$ and $\mathrm{Nd^{3+}}$ of the empirically determined parameters in the conventional Hamiltonian of the form

$$\begin{aligned}\beta\mathbf{H}\,.\,(\mathbf{L}+2\mathbf{S}) &+ hcA_2^0(3z^2-r^2) + hcA_4^0(35z^4-30r^2z^2+3r^4)\\ &+ hcA_6^0(231z^6-315r^2z^4+105r^4z^2-5r^6)\\ &+ hcA_6^6(x^6-15x^4y^2+15x^2y^4-y^6). \qquad \ldots\ldots(2)\end{aligned}$$

In this Hamiltonian, applicable to ions with non-zero orbital angular momentum, the A coefficients are determined in a relatively direct manner by the position and charges of the other ions in the crystal, whereas the analogous constants B_n^m in (1) depend on the crystal field in a much more complicated manner, through whatever mechanism is operative in producing the splitting of the S state.

§ 5. DISCUSSION OF RESULTS

Various mechanisms have been proposed to account for the splitting of S states by crystalline electric fields, and the results of the present experiments are best discussed in terms of these models. A detailed theoretical examination is at present in progress, and the results that are quoted below without derivation will be dealt with at greater length in a later paper (Judd, Pryce and Stevens, to be published). The aim of this discussion is to provide a picture of the situation in semi-quantitative terms.

The present experimental results display the feature mentioned by Elliott and Stevens (1953), namely, that b_2^0 is much larger than the other coefficients, whereas the effect of $A_2^0\langle r^2\rangle$ is usually smaller than that of the other parameters

Table 2. Comparison of Parameters

		Ethylsulphate		Double Nitrate		Trichloride	
		90°K a	290°K b	90°K c	290°K	90°K d	290°K d
Gd^{3+}	b_2^0	$+(204{\cdot}7 \pm 2\ ) \times 10^{-4}$	$+191 \times 10^{-4}$	$+124 \times 10^{-4}$	—	$+(16{\cdot}0 \pm 0{\cdot}2\ ) \times 10^{-4}$	$(8{\cdot}36 \pm 0{\cdot}10) \times 10^{-4}$
	b_4^0	$-(3{\cdot}96 \pm 0{\cdot}3) \times 10^{-4}$	$-3{\cdot}68 \times 10^{-4}$	$+0{\cdot}9 \times 10^{-4}$	—	$+(2$[illegible]$3 \pm 0{\cdot}05) \times 10^{-4}$	$(1{\cdot}68 \pm 0{\cdot}04) \times 10^{-4}$
	b_6^0	$+(0{\cdot}63 \pm 0{\cdot}1) \times 10^{-4}$	$0{\cdot}53 \times 10^{-4}$	$+0{\cdot}6 \times 10^{-4}$	—	$+(0{\cdot}25 \pm 0{\cdot}05) \times 10^{-4}$	$(0{\cdot}64 \pm 0{\cdot}15) \times 10^{-4}$
	b_6^6	$-(3{\cdot}5 \pm 5\ ) \times 10^{-4}$	$27{\cdot}0 \times 10^{-4}$	12×10^{-4}	—	$+(1{\cdot}4 \pm 0{\cdot}3\ ) \times 10^{-4}$	—
Nd^{3+}	$A_2^0 \langle r^2 \rangle$	-15	e	-50	c	—	f
	$A_4^0 \langle r^4 \rangle$	-35		-30		—	
	$A_6^0 \langle r^6 \rangle$	-60		-40		—	
	$A_6^6 \langle r^6 \rangle$	$+640$		$+700$		—	
Pr^{3+}	$A_2^0 \langle r^2 \rangle$	-15	g	-70	c	$+61$	h
	$A_4^0 \langle r^4 \rangle$	-38		-20		-46	
	$A_6^0 \langle r^6 \rangle$	-76		-50		-35	
	$A_6^6 \langle r^6 \rangle$	$+895$		$+700$		355	

All values are in cm^{-1}

a, Bleaney *et al.* (1954); b, Scovil (1951); c, Judd (1955 b); d, present paper; e, Elliott and Stevens (1953); f, the ratio of the A_6^0 and A_6^6 calculated by Hutchison, Wong and Dorain (to be published) gives g values in good agreement with experiment for Nd^{3+}; g, these numbers result from a linear interpolation between the values for Ce^{3+} as given by Elliott and Stevens (1952) and those given for Nd^{3+} by Elliott and Stevens (1953); h, these values are calculated from the optical data of Sayre, Sancier and Freed (1955) and the theory of Elliott and Stevens (1953) using the ratio A_6^0/A_6^6 calculated by Hutchison, Wong and Dorain.

in the cases of rare-earth ions not in S states. They point out that there are two matrix elements of the crystal field V for which the terms in $A_4^0\langle r^4\rangle$, $A_6^0\langle r^6\rangle$ and $A_6^6\langle r^6\rangle$ vanish, but the one in $A_2^0\langle r^2\rangle$ does not. These are

(*a*) $$\langle 4f^7, {}^6D_{7/2}|V|4f^7, {}^6P_{7/2}\rangle$$

and

(*b*) $$\langle 4f^6 6p, {}^8D_{7/2}|V|4f^7, {}^8S_{7/2}\rangle.$$

The first may be used within the configuration $4f^7$ together with the spin–orbit interaction Λ to give a splitting of the ground state by means of a fourth-order mechanism which may be schematically represented as follows:

(i) $$\langle {}^8S_{7/2}|\Lambda|{}^6P_{7/2}\rangle\langle {}^6P_{7/2}|\Lambda|{}^6D_{7/2}\rangle\langle {}^6D_{7/2}|V|{}^6P_{7/2}\rangle\langle {}^6P_{7/2}|\Lambda|{}^8S_{7/2}\rangle.$$

The second, as suggested by Pryce (1950), may be used together with the spin–spin interaction V_{ss} to give a splitting by means of the following second-order mechanism:

(ii) $$\langle 4f^7, {}^8S_{7/2}|V_{ss}|4f^6 6p, {}^8D_{7/2}\rangle\langle 4f^6 6p, {}^8D_{7/2}|V|4f^7, {}^8S_{7/2}\rangle.$$

The ${}^8D_{7/2}$ levels in other excited configurations of odd parity will also contribute; $4f^6 6p$ is chosen because it is probably the lowest of such configurations.

Both of these mechanisms will lead to a b_2^0 term only in the spin-Hamiltonian because higher-order terms in V vanish, and both are linear in the crystalline electric field.

The fact that both (i) and (ii) are mechanisms which give an effect linear in the crystalline field enables one to discard them as the sole mechanisms by means of which the ground state of Gd^{3+} is split. This is seen to be the case on inspection of table 2, where no proportionality between the values of b_2^0 and $A_2^0\langle r^2\rangle$ in the different crystals is found.

Another mechanism has been proposed by Judd (1955 a), in which the crystalline field perturbation together with the spin–orbit interaction results in a splitting of the ground state through a fourth-order mechanism which may be schematically represented as follows:

(iii) $$\langle {}^8S_{7/2}|\Lambda|{}^6P_{7/2}\rangle\langle {}^6P_{7/2}|V|{}^6X_J\rangle\langle {}^6X_J|V|{}^6P_{7/2}\rangle\langle {}^6P_{7/2}|\Lambda|{}^8S_{7/2}\rangle.$$

In this mechanism an intermediate P state in the same configuration f^7 is split by the crystalline field interaction with higher states ${}^6X_J$. Then, since the P state interacts with the ground state through spin–orbit coupling, the splitting of the P state is reflected in a splitting of the ground S state.

In the first place it should be noted that terms of all orders in the crystalline field will contribute to mechanism (iii) through appropriate states, ${}^6X_J$. However, because of the nature of the intermediate state, this mechanism will lead only to a term b_2^0 for the spin-Hamiltonian (1). Moreover, since the sixth-order terms in the crystalline field potential V are the largest for the rare earths (Elliott and Stevens 1953), the size of the term in b_2^0 will be determined chiefly by the size of the sixth-order terms in V. It is to be noted that the states ${}^6X_J$ in configurations of even parity will also contribute to b_2^0 through parameters such as A_5^3.

In the second place, it should be observed that this mechanism is quadratic in V and involves a large number of elements $\langle {}^6P_{7/2}|V|{}^6X_J\rangle$ for a variety of states ${}^6X_J$. Consequently, there is no expectation of proportionality or any other simple relation between the values of b and of $A_n^m\langle r^n\rangle$ in various crystals. The model is in this respect consistent with the observations.

Two other facts which argue effectively for the importance of mechanism (iii) are:

(*a*) The value of $A_6^6\langle r^6\rangle$ and also of $A_6^0\langle r^6\rangle$ for the ethylsulphate is from 2·2 to 2·5 times as large as for the trichloride in the case of Pr^{3+}. Since these terms should be two of the most important in producing the splitting in Gd^{3+}, we would expect b_2^0 for the ethylsulphate to be from $2{\cdot}2^2$ to $2{\cdot}5^2$ (∼5 to 6) times as large as b_2^0 for the trichloride. The ratio is actually considerably larger than about 5, and this may be accounted for when mechanism (i) is taken into consideration and assumed to be acting along with (iii). This will be discussed below.

(*b*) Since mechanism (iii) is quadratic in V, and since the sixth-order terms in V (which depend on the inverse seventh power of the inter-ionic distances in the crystal) are predominant in V, we may expect the splitting to vary as the inverse fourteenth power of the lattice spacings in a given crystal. If we assume the linear coefficient of thermal expansion for the crystal to be $30\times10^{-6}\,\mathrm{deg}^{-1}$, we would expect the splitting to change by a factor 1·09 between 90°K and 290°K and to be smaller at the higher temperature. The observed b_2^0 is smaller at the higher temperature by the factor 1·07 in the case of the ethylsulphate and by the very large factor 1·9 in the case of the trichloride. Again, the unexpectedly large factor in the latter case may be explained when mechanism (i) is taken into account. In the case of either mechanism (i) or (ii) the value of the constant b_2^0 will depend on the third power of the lattice spacings and so may be expected to give a splitting that is relatively independent of the temperature.

The contribution to b_2^0 from (i) can be shown to be

$$-12\zeta^2 A_2^0\langle r^2\rangle/5W_P^2W_D,$$

where W_P and W_D are the energies of $^6P_{7/2}$ and $^6D_{7/2}$ relative to $^8S_{7/2}$, and ζ is the spin–orbit coupling constant. These quantities can be estimated from Judd (1956), and the final result is that for a value of 30 cm^{-1} for $A_2^0\langle r^2\rangle$, which seems reasonable for $GdCl_3$, the contribution to b_2^0 is $-75\times10^{-4}\,\mathrm{cm}^{-1}$. Detailed calculations for mechanism (ii) are more difficult, since excited configurations are involved; it seems unlikely, however, that it is as important as (i).

Since $A_2^0\langle r^2\rangle$ for gadolinium ethylsulphate is extremely small (Elliott and Stevens (1953) give $A_2^0\langle r^2\rangle=0$ for all diluted rare-earth ethylsulphates beyond Nd^{3+}), we may deduce that the major contribution to b_2^0 comes from the quadratic effect. Dividing the observed value of $205\times10^{-4}\,\mathrm{cm}^{-1}$ by five, for reasons given above in paragraph (*b*), the value of $+41\times10^{-4}\,\mathrm{cm}^{-1}$ is obtained for the quadratic effect in gadolinium trichloride. Thus we see that the contributions of mechanisms (i) and (iii) are of the same order of magnitude and opposite in sign. The fact that the observed value for b_2^0 is $+16\times10^{-4}\,\mathrm{cm}^{-1}$ at 90°K suggests that the quadratic effect (iii) slightly predominates over (i) at this temperature. This is not at all surprising, as other components of the crystalline field contribute to (iii), and the value of 41×10^{-4} deduced above must be regarded as an approximate one.

Another test of the consistency of the theory is provided by the temperature dependence of b_2^0. We have already noted that a linear coefficient of thermal expansion of $30\times10^{-6}\,\mathrm{deg}^{-1}$ leads to a 9% reduction in the mechanism (iii) in passing from 90°K to room temperature. In the case of the chloride, such a reduction would be approximately $7 \times 10^{-4}\,\mathrm{cm}^{-1}$ for a quadratic effect of $80 \times 10^{-4}\,\mathrm{cm}^{-1}$; and since mechanism (iii) is positive, we would expect the observed value of b_2^0 at room temperature to be about $(16-7) \times 10^{-4}\,\mathrm{cm}^{-1}$, i.e. 9×10^{-4}. The value actually observed is $8{\cdot}36 \times 10^{-4}$ (see table 1). In this calculation no account has been taken of the non-uniform contraction of the crystal, since little is known about this effect at present.

This calculation has been necessarily very approximate, since the relevant parameters are not yet known with accuracy; what we have tried to show is that the theory is entirely consistent with the experimental results to within an order of magnitude.

ACKNOWLEDGMENTS

Dr. P. M. Llewellyn constructed the apparatus with which the experiments reported in this paper were performed and was of invaluable assistance in their execution. We thank Professor M. H. L. Pryce for helpful conversations concerning this work and also Eugene Wong for the preparation of $CeCl_3$ and $LaCl_3$. One of us (C. A. H.) is indebted to the Clarendon Laboratory for the privilege of visiting and working there.

REFERENCES

ABRAGAM, A., and PRYCE, M. H. L., 1951, *Proc. Roy. Soc.* A, **205,** 135.

ANDERSON, J. H., and HUTCHISON, C. A., JR., 1955, *Phys. Rev.*, **97,** 76.

BLEANEY, B., ELLIOTT, R. J., SCOVIL, H. E. D., and TRENAM, R. S., 1951, *Phil. Mag.*, **42,** 1062.

BLEANEY, B., SCOVIL, H. E. D., and TRENAM, R. S., 1954, *Proc. Roy. Soc.* A, **223,** 15.

BLEANEY, B., HUTCHISON, C. A., JR., LLEWELLYN, P. M., and POPE, D. F. D., 1956, *Proc. Phys. Soc.* B, **69,** 1167.

DORAIN, P., HUTCHISON, C. A., JR., and WONG, E., 1957, *Phys. Rev.*, in the press.

ELLIOTT, R. J., and STEVENS, K. W. H., 1952, *Proc. Roy. Soc.* A, **215,** 437; 1953, *Ibid.*, **219,** 387.

HUTCHISON, C. A., JR., LLEWELLYN, P. M., WONG, E., and DORAIN, P., 1956, *Phys. Rev.*, **102,** 292.

JUDD, B. R., 1955 a, *Thesis*, University of Oxford; 1955 b, *Proc. Roy. Soc.* A, **232,** 458; 1956, *Proc. Phys. Soc.* A, **69,** 157.

PRYCE, M. H. L., 1950, *Phys. Rev.*, **80,** 1106.

SAYRE, E. V., SANCIER, K. M., and FREED, S., 1955, *J. Chem. Phys.*, **23,** 2060.

SCOVIL, H. E. D., 1951, *Thesis*, University of Oxford.

TRENAM, R. S., 1953, *Proc. Phys. Soc.* A, **66,** 118.

VAN VLECK, J. H., and PENNEY, W. G., 1934, *Phil. Mag.*, **17,** 961.

ZACHARIASEN, W. H., 1948, *J. Chem. Phys.*, **16,** 254.

Induction of Ring Cracks on Diamond Surfaces

BY S. TOLANSKY AND V. R. HOWES
Royal Holloway College (University of London), Egham, Surrey

MS. received 23*rd November* 1956

Abstract. In extension of earlier studies on the production of percussion marks on diamond surface using a diamond sphere as indenter, it is now found that much softer materials than diamond can be used to produce pressure crack figures on diamond surfaces. Tungsten carbide and sapphire balls have produced cracking on an octahedral face of diamond, this face having been already shown to be the face of least resistance to this type of deformation. The cracks are of the same nature as those previously produced by a diamond indenter and the critical stress necessary is of the same order, but the cracks are on a bigger scale and have greater associated disturbance. The full development of a pressure crack is traced and multiple beam interference pictures show the extent of the surface damage. Three distinct types of cracking are found associated with each percussion mark and the mechanism of each is discussed in terms of cleavage, crystallographic shatter and cracks due to shock-wave propagation.

---

§ 1. EXPERIMENTAL

THE production of oriented pressure cracks on diamond faces which have been subject to pressure with a diamond sphere has already been discussed (Howes and Tolansky 1955). The cracks produced artificially closely resemble those found by careful microscopic examination on many good quality South African diamonds obtained through deep mining operations. It was conjectured that these naturally-occurring cracks might be produced in the ore-crushing processes employed in production. Experiments have therefore been carried out on diamond surfaces using ball impactors of materials considerably softer than diamond. It is shown here that they too induce ring cracks. Balls of tungsten carbide and of synthetic sapphire produced clear pressure-cracks on the octahedral face of a diamond, and in the former case considerable internal disruption of the diamond was observed. No cracking could be obtained on either cubic or dodecahedral faces. In all tests, the balls themselves have been severely damaged.

The spheres, of 1 mm diameter, were fitted in a hardness testing instrument mounted on to the stage of an inverted type projection microscope. The sphere was pressed down on to a parallel-sided transparent diamond specimen and observation made on the region of contact by the microscope system immediately underneath the specimen whilst load was maintained. This arrangement allowed much higher magnifications to be secured than in former experiments, and the development of the cracking could be pursued whilst pressure was varied. The load was applied such that a regular slow rate of increase was possible. Three diamond surfaces were examined, namely, a natural octahedral face of a parallel-sided transparent diamond portrait stone, an approximately dodecahedral plane of a diamond of octahedral habit, and an approximately cubic plane of another diamond of octahedral habit. These last two stones were both in the form of transparent parallel-sided plates, obtained by sawing and polishing octahedral stones appropriately.

Loads were applied up to a maximum of 30 kg, and after the test the specimens were examined by the microscope with transmission polarized illumination, to study the disruption within the body of the crystal. The damaged diamond surface was then silvered, and the surface distortions studied, by multiple-beam Fizeau fringes and fringes of equal chromatic order (Tolansky 1948) and by phase-contrast.

Despite the fact that tungsten carbide is considerably softer than diamond a ring crack develops with a carbide ball and the evaluation of this is shown (× 235) in figure 1 (*a*) to (*e*), (Plate I). Figure (*a*), with zero load, shows Newton's rings. Figure 1 (*b*), with 3 kg load shows as black spot the area of true contact, surrounded by Newton's rings. At 1(*c*), load 16 kg, the ring crack abruptly develops. It is in shape a crystallographically oriented hexagon which almost exactly circumscribes the circle of contact, although not quite. The oriented ring crack has three symmetrically placed surrounding dark areas (which we will call ***simple*** internal cleavage fractures) and markedly internally disrupted regions (which we call ***complex*** internal fractures). As the load was further increased both of these internal effects developed—but not symmetrically. For a load of 20 kg figure 1(*d*), the shadow area at the top is seen to have extended before the other two, and in 1(*e*) (30 kg) two areas have developed while the third has been inhibited. This last photograph was for the maximum load of 30 kg given by our machine, and presumably if it had been possible to increase the load further, the third dark area would eventually have spread out to resemble the other two areas. In the hope that this might still happen the 30 kg load was left applied for an hour, by which time the value had decreased to 28·8 kg (probably due to relaxation of the apparatus under the extreme stress conditions). The load was again increased up to 30 kg but no further development occurred.

Slow reduction of load was begun, and at 29·2 kg the third shadow area suddenly extended to give the effectively symmetrical pattern seen in figure 1(*f*). Further reduction of load to zero was then carried out. Whereas the ***simple*** internal shadow areas retreat progressively as the load is decreased, figure 1(*g*) the ***complex*** disturbances opposite the other three sides remain unchanged in appearance, figure 1(*h*). The shadow areas retreated asymmetrically, the first to disappear being the last to develop fully.

§ 2. Optical Study of the Pressure Figure

The dark shadow areas (formerly observed also while the specimen was under load) have been explained previously as simple internal cleavage cracks along (111) planes in the crystal. Since these simple internal cracks disappear when the load is removed, they can only be observed during load. The complex internal fractures, examined in transmission with polarized light, are shown in figure 1(*i*) (× 300). They occur below the surface, the outline of the surface hexagon figure being out of focus. The deeper below the surface, the further are these complex cracks away from the rectilinear hexagonal pressure mark, which indicates that they occur roughly on planes at an acute angle to the surface, going away from the pressure figure.

Such a photograph reveals that there are two components in the complex fracture. The details are not in focus in any one picture. The first is seen to be sets of multiple straight lines, there being two sets for each of the three alternate sides of the hexagon as shown schematically in figure 2. These can also be seen clearly in the photographs of the series taken while under load. There are three principal directions for these lines which are parallel to the *other* three alternate sides of the hexagon, i.e. to the sides where the shadow areas occur while under load. Thus the three directions are identical with the directions of the lines of intersection of crystallographic (111) planes with the octahedral face.

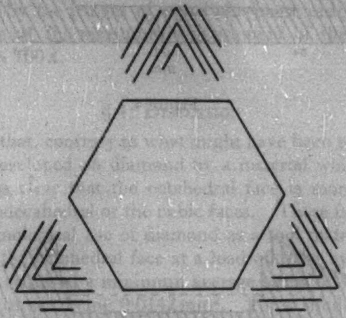


Figure 2.

The second type of secondary internal disturbance is clearly seen in focus in figure 1(*i*), and lies beyond the first type—also being deeper in the crystal. This occurs opposite the same three alternate sides of the hexagon as above, and consists of three sets of slightly curving irregular cracks running approximately parallel to the sides. This phenomenon is exactly similar to that observed previously for a multiple pressure crack formed on an octahedral face by a diamond ball, but is more distinct and extensive; it also corresponds with the secondary internal cracks found associated with the pressure figure formed on an approximate cubic plane by a diamond ball (Howes and Tolansky 1955), in that the lines are curving and non-crystallographic.

The damaged surface was silvered and figure 1(*j*) (× 550) shows the reflection, phase-contrast photograph of the surface hexagonal crack. It is indeed complex. The primary crack is an almost perfect hexagon while the outer crack of the multiple figure is very imperfect in its hexagonal shape, although it is still approximately orientated.

The silvered surface of the diamond was then matched against a silvered glass flat in order to obtain multiple-beam interference fringes, for the detection of the surface distortion accompanying the formation of the pressure crack. The Fizeau fringe pattern in figure 3(*a*) (Plate II) (× 200) shows a fringe bisecting opposite sides of the hexagonal figure. Quantitative measurements are made using the fringes of equal chromatic order (figures 3(*b*) and 3(*c*) (× 300)), and the calibrated sectional profiles given by fringes bisecting opposite sides and opposite angles of the figure are drawn to scale in figures 4 and 5 where the insets show the direction of the profile in each case. The surface distortion, as with the internal damage, is on a larger scale than that previously found with the diamond ball experiments, the height displacements being twice as large.

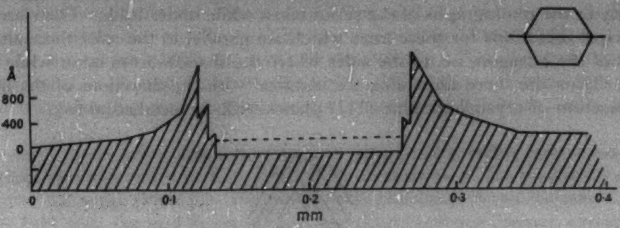


Figure 4.

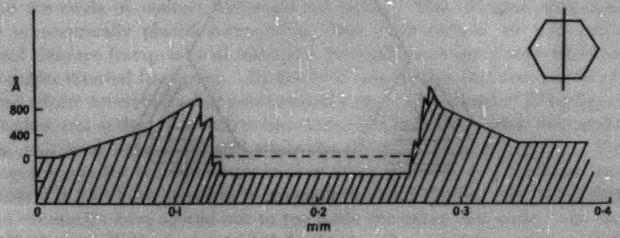


Figure 5.

After the application of the 30 kg load the tungsten carbide balls were found to be appreciably flattened.

With tungsten carbide balls of approximately 1 mm diameter, tests were then made on dodecahedral and cubic planes up to the full maximum load of 30 kg without any cracking being obtained. After both of these tests the surfaces were examined at the region where contact had been made, but no change in the surfaces could be detected interferometrically at all. The balls, in both tests, were also considerably flattened at their tips.

With the deformation of the balls, the applied stress is reduced, for the same load, since the area of contact increases. The maximum load in all tests was 30 kg, and by observing through the crystal the corresponding area of contact was seen to be the same on both dodecahedral and cubic planes, whence for both surfaces, the maximum applied average stress is $1{\cdot}2 \times 10^{11}$ dyn cm^{-2}. This compares with a real average stress of $1{\cdot}1 \times 10^{11}$ dyn cm^{-2}, necessary for initiation of cracking on the octahedral face by the tungsten carbide 1 mm ball. Thus it is established that the octahedral face is less resistant to pressure cracking than the other two planes: it is also to be noted that in spite of an increase in load of from 16 kg up to 30 kg, the resulting stress increased relatively little owing to the associated increased deformation of the ball.

§ 3. SAPPHIRE BALL

A synthetic sapphire ball of approximately 1 mm diameter was applied to the octahedral face of a diamond, this being the less resistant face. The load was steadily increased to a maximum of 9·6 kg at which the tip of the sapphire shattered. The load was increased no further for fear of a possible complete disintegration under high stress; even so, a partly developed true pressure ring crack was formed on the diamond face figure 3(*d*).

This was clearly seen after the removal of load. The figure is more rounded than has been seen previously, this being most likely due to the deformation of the sapphire before breakage and thus to an alteration of the stress concentration within the diamond.

That the surface distortion for this crack figure is of the same general pattern is shown in figure 3 (*e*) by the fringes of equal chromatic order but the height displacement is very small, the maximum displacement of the pile-up around the crack being less than 100 Å.

§ 4. DISCUSSION

It has been seen that, contrary to what might have been expected, an oriented ring crack can be developed on diamond by a material which is not as hard as diamond itself: it is clear that the octahedral face is more vulnerable to this cracking than the dodecahedral or the cubic faces. These facts are of interest in the field of the technological use of diamond as a tool material. The pressure figure produced on the octahedral face at a load of 16 kg by the 1 mm diameter *tungsten carbide* ball required a minimum average stress of $1{\cdot}1 \times 10^{11}$ dyn cm^{-2}—the critical area of contact being 0·0142 mm^2. For a similar octahedral face, a *diamond* ball impactor of 0·78 mm diameter produced cracking at an average stress of $1{\cdot}4 \times 10^{11}$ dyn cm^{-2} for both the primary pressure figure and the simple internal fracture of a multiple figure, the critical areas of contact being 0·0036 and 0·0025 mm^2 (loads 5 kg and 3·5 kg respectively). It would appear from a comparison of the above data that, as the scale of the experiment is increased, the strength of the surface decreases. This may possibly be due to a strength–area relationship similar to that found for glass, local weak areas being there the cause. Alternatively the permanent deformation of the carbide ball may lead to an altered stress distribution. Although the stress necessary to initiate the cracking in these experiments was less than in the previous tests, the complex internal fractures and also the surface pile-up was considerably more. This is explainable by taking into account a greater energy release once cracking is initiated, due to the release of the potential energy stored in the elastically deformed tungsten carbide ball. It is because the diamond ball is much more resistant to elastic deformation that cracking in the former tests occurred at considerably lower loads—the stress being already high, due to the small areas of contact, through lack of deformation.

The pressure figure formed on the diamond surface with the sapphire impactor (although of the same diameter as the tungsten carbide) was much smaller (being 55×10^{-6} cm^2 compared with 261×10^{-6} cm^2), but it was not possible to measure the critical area of contact nor the critical load because of the superimposed breaking up pattern of the sapphire. The pressure figure obtained was formed at a load of 9·6 kg so that it would appear that less elastic deformation occurred than with the tungsten carbide ball. Further, the diamond shows no internal disturbance associated with this crack, indicating a minimum of energy release at cracking. This was to be expected since cracking of the diamond occurred *after* the fracturing of the sapphire ball itself, which would absorb stored-up elastic-energy.

REFERENCES

HOWES, V. R., and TOLANSKY, S., 1955, *Proc. Roy. Soc.* A, 230, 287 (Parts I and II).
TOLANSKY, S., 1948, *Multiple Beam Interferometry* (Oxford : Clarendon Press).

RESEARCH NOTES

Magnetoresistance of a p-type Semiconducting Diamond

By E. W. J. MITCHELL and P. T. WEDEPOHL
Physics Research Laboratories, University of Reading

Communicated by R. W. Ditchburn; MS. received 25th January 1957

§ 1. Introduction

Measurements of the semiconducting properties of type IIb diamonds have been reported by several workers, e.g. Custers (1955), Austin and Wolfe (1956), Dyer and Wedepohl (1956) and Wedepohl (1957). A transverse magnetoresistance was briefly reported by Wedepohl (1957), and in the present note the results of more detailed measurements are described.

§ 2. Experimental

The crystal used (D111) was a semiconducting p-type rectangular diamond block whose dimensions were 0·401 cm × 0·191 cm × 0·174 cm. At 290°K (the temperature at which the measurements were made), the Hall coefficient R was $4{\cdot}3\times10^5$ cm^3 coulomb^{-1}, and the resistivity ρ 300 ohm cm. The crystallographic orientations of the length, width and thickness axes of the crystal were determined from a back-reflection Laue photograph. Relative to the cube axes, the direction cosines of the specimen axes are as follows:

	l	m	n
Length axis	−0·618	−0·02	+0·792
Width axis	+0·19	−0·968	+0·12
Thickness axis	+0·766	+0·26	+0·593

The angles from which the direction cosines were determined are subject to a probable error of about ±1°. This leads to a large uncertainty in the smaller numerical values of l, m or n, but the small values are not important in the analysis of the results (cf. equation (2)).

In the experiment the magnetic field was horizontal and the specimen was mounted so that it could be rotated about a vertical direction, which was either the width axis or the thickness axis. The current direction was always along the length of the specimen. Potential probes were used to measure the voltage drop along the specimen. The current was also measured potentiometrically.

Since the resistance of the diamond is highly dependent on temperature, strict precautions were taken to reduce temperature variations as far as possible, and to correct for their effect where necessary. This was done by 'saddling' each measurement of resistance in the magnetic field by two zero-field measurements and using the mean of these two when calculating the change in resistance due to the applied field.

§ 3. Results

3.1. *Magnetoresistance as a Function of Field Strength*

The field dependence of the magnetoresistance was investigated for the case of magnetic field parallel to the width axis (transverse effect). It was found that the value of $\Delta r/r_0$ (where Δr is the resistance change and r_0 the zero-field resistance) was altered when the magnetic field was reversed. This effect was due to slight misalignment of the probes, giving rise to a small component of Hall voltage between them. The difference between the forward and reverse values of $\Delta r/r_0$ was proportional to the first power of the magnetic field, as would be expected if a Hall component were the cause.

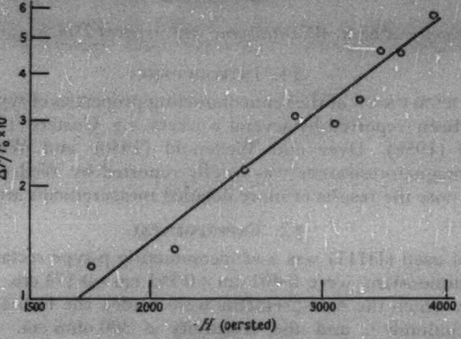


Figure 1. Proportional change in resistance as a function of magnetic field. The straight line represents a square law.

The effect of such a misalignment of the probes may be eliminated from the results by taking the mean value of $\Delta r/r_0$ for forward and reverse magnetic fields. In figure 1 the mean values of $\Delta r/r_0$ are plotted against H, the scales being logarithmic. The relation $\Delta r/r_0 = 3{\cdot}54 \times 10^{-10} H^2$ is also plotted and is in reasonable agreement with the experimental points. This is in accordance with existing low-field magnetoresistance theory.

3.2. *Magnetoresistance as a Function of Angle between Current and Magnetic Field Vectors*

In figure 2, the mean value of $\Delta r/r_0$ for forward and reverse fields is plotted against θ, which is the angle between the current vector **i** and the magnetic field

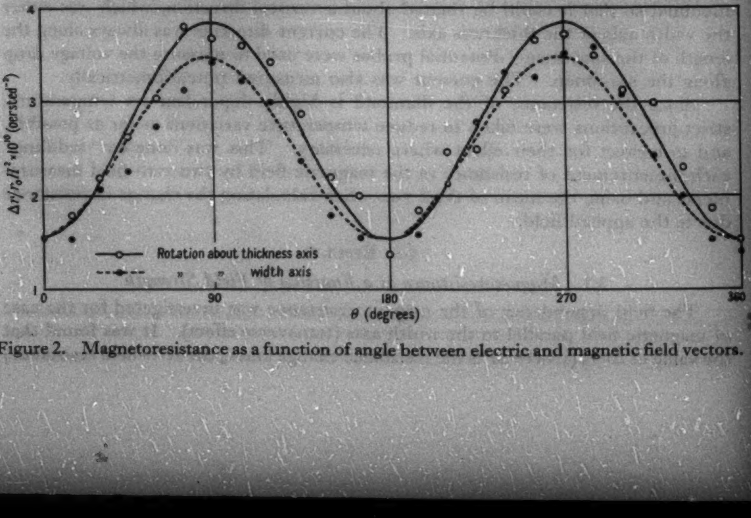


Figure 2. Magnetoresistance as a function of angle between electric and magnetic field vectors.

vector $\mathbf{H}$. The two plots are very similar and show that the effect is a maximum for $\mathbf{H}$ perpendicular to $\mathbf{i}$ and a minimum for $\mathbf{H}$ parallel to $\mathbf{i}$. The field used was 3630 oersteds.

§4. DISCUSSION OF RESULTS

Wick (1954) has shown that a correction must be applied to magnetoresistance measurements made on short samples, due to the fact that the equipotential planes are distorted near the ends of the specimen when a magnetic field is applied. He has treated the case of transverse magnetoresistance in samples having perfectly conducting current electrodes, the resistance being measured between these electrodes. For this case the measured value of $\Delta r/r_0H^2$ is too large by an amount

$$\left(\frac{\Delta r}{r_0H^2}\right)' = 6{\cdot}37 \times 10^{-17}\,\frac{W}{L}\left(\frac{R}{\rho}\right)^2 \qquad \ldots\ldots(1)$$

where W is the width of the sample perpendicular to the field and L is its length. This correction would have been $0{\cdot}6 \times 10^{-10}$ oersted^{-2} for the diamond used (cf. figure 2). The case of potential probes has not been treated by Wick, but the correction is probably smaller for this case. To ensure good contact the probes were placed on an edge of the diamond and were symmetrical about the centre. The probe separation was 0·205 cm for the case of rotation about the thickness axis, and 0·167 cm for rotation about the width axis.

The Wick correction is a maximum for the transverse effect, and is zero for the longitudinal effect.

A theoretical treatment of magnetoresistance in semiconductors, due to Seitz, leads to an equation of the form

$$\frac{\Delta r}{r_0H^2} = b + c\cos^2\theta + d(l_1^2l_2^2 + m_1^2m_2^2 + n_1^2n_2^2) \qquad \ldots\ldots(2)$$

(see for example Pearson and Suhl 1951). Here l_1, m_1, n_1 and l_2, m_2, n_2 are the direction cosines of $\mathbf{i}$ and $\mathbf{H}$ respectively, relative to the major crystallographic axes, and b, c, d are constants of the specimen for a given temperature. This treatment is valid at low fields (those for which $\Delta r/r_0$ varies as H^2) for any semiconductor possessing cubic symmetry.

In our case d was determined from each pair of results at the same θ (see equation (2)). Values of b and c were found by using the mean value of d in equation (2), and suitably combining the experimental results in pairs. The mean values of b, c and d, together with the r.m.s. errors of the means, are

$$b = (+3{\cdot}87 \pm 0{\cdot}06) \times 10^{-10}\ \mathrm{Oe}^{-2},$$
$$c = (-1{\cdot}83 \pm 0{\cdot}12) \times 10^{-10}\ \mathrm{Oe}^{-2},$$
$$d = (-0{\cdot}91 \pm 0{\cdot}23) \times 10^{-10}\ \mathrm{Oe}^{-2}.$$

Curves calculated from equation (2), using these values of b, c and d, are shown in figure 2. The full curve corresponds to rotation about the thickness axis, the broken curve to rotation about the width axis.

§5. CONCLUSION

If the experimental values of b, c and d are substituted in equation (2), it is found that (*a*) although $\Delta r/r_0H^2$ has a pronounced dependence upon θ it is not very sensitive to the orientation of $\mathbf{i}$ and $\mathbf{H}$ relative to the crystallographic axes, and (*b*) the longitudinal magnetoresistance is smaller than the transverse for all directions.

These features are also found in the magnetoresistance measurements on p-type germanium by Pearson and Suhl (1951), and on p-type silicon by Pearson and Herring (1954). The corresponding work done by the same workers on n-type germanium and silicon gives quite different results, which show a strong dependence of the magnetoresistance on the crystallographic orientation of **i** and **H** and a longitudinal magnetoresistance larger than the transverse in some cases.

Our results are inadequate for an unambiguous deduction of the band structure. However, the low longitudinal magnetoresistance in all directions is consistent with a model in which the band edge occurs at the centre of the Brillouin zone, and in which each surface of constant energy consists of one or more approximately spherical shells which touch at the band edge (e.g. Kittel 1954). There is strong evidence to support this model in the case of silicon and germanium from cyclotron resonance work (Dresselhaus *et al.* 1954, Lax *et al.* 1954).

ACKNOWLEDGMENTS

We should like to thank Messrs. Industrial Distributors (1946) Ltd. for supporting this research, Dr. J. F. H. Custers, Director of Research, Diamond Research Laboratory, Johannesburg, for supplying the specimen used in the investigation, Professor R. W. Ditchburn for laboratory facilities, and Drs. R. J. Elliott, J. Drabble and R. Wolfe for valuable discussions.

REFERENCES

AUSTIN, I. G., and WOLFE, R., 1956, *Proc. Phys. Soc.* B, **69**, 329.
CUSTERS, J. F. H., 1955, *Nature, Lond.*, **176**, 173.
DRESSELHAUS, G., KIP, A. F., and KITTEL, C., 1954, *Phys, Rev.*, **95**, 568.
DYER, H. B., and WEDEPOHL, P. T., 1956, *Proc. Phys. Soc.* B, **69**, 410.
KITTEL, C., 1954, *Physica*, **20**, 829.
LAX, B., ZEIGER, H. J., and DEXTER, R. N., 1954, *Physica*, **20**, 818.
PEARSON, G. L., and HERRING, C., 1954, *Physica*, **20**, 975.†
PEARSON, G. L., and SUHL, H., 1951, *Phys. Rev.*, **83**, 768.‡
WEDEPOHL, P. T., 1957, *Proc. Phys. Soc.* B, **70**, 177.
WICK, R. F., 1954, *J. Appl. Phys.*, **25**, 741.

† The horizontal axis of figure 2 in this paper should be displaced along its length by 90° (Pearson, G. L., private communication).

‡ The captions to figures 8 and 9 in this paper should be interchanged.

The Magnetoresistance of the Nickel–Zinc Ferrite System

By R. PARKER
Department of Physics, University College of North Wales, Bangor

MS. received 30th January 1957

The magnetoresistance of ferromagnetic metals and alloys has been studied extensively and the empirical laws governing this phenomenon have been firmly established. There is, however, no corresponding information available regarding the behaviour of semiconductors which exhibit magnetic ordering below their Curie temperature. Magnetoresistance measurements have been reported only for magnetite (Masumoto and Shirakawa 1941, Domenicali 1950) and for certain manganites at low temperatures (Volger 1954). These materials have relatively low electrical resistivities and exhibit quasi-metallic conductivity over certain temperature ranges (Smith 1952, van Santen and Jonker 1950).

The present investigation was undertaken with a view to establishing the laws governing the magnetoresistance of semiconducting ferrites, which have high specific resistivities at room temperature. D.C. measurements have so far been made with a number of specimens in the nickel–zinc ferrite system in the composition range $Ni_{(1-x)}Zn_xFe_2O_4$, with x varying from 0 to 0·8; the main results obtained are briefly summarized in this note.

Specimens employed were fired between 1160° and 1300°c and had room temperature resistivities of approximately 10^8 ohm cm. The conductivity, as judged from thermoelectric measurements was found to be p-type in every case. The temperature variation of resistivity above 50°c was found to be given by the relationship $\rho = C \exp(\epsilon/\boldsymbol{k}T)$, the constant ϵ taking values between 0·3–0·5 ev. At lower temperatures the same functional relationship was observed with somewhat smaller values for ϵ. These results are broadly in agreement with those reported by van Uitert (1956).

The slope of the resistivity–temperature curves of ferromagnetic metals and alloys shows a marked discontinuity in the region of the magnetic transformation temperature, the actual resistivity below the Curie point being much lower than that extrapolated from the paramagnetic region. Analogous discontinuities have been observed in the relationship between $\log \rho$ and $1/T$ for magnetite and some of the manganites. A search was made for a similar effect in the specimens under investigation. The Curie temperature of the $Ni_{(1-x)}Zn_xFe_2O_4$ system decreases from about 600°c for $x=0$ to about 80°c for $x=0{\cdot}8$ (Guillaud 1950). No anomaly was found in the $(\log \rho, 1/T)$ curves at the Curie temperature for any specimen in this composition range. It may be concluded therefore that the conduction process in nickel–zinc ferrites is not affected by the advent of spontaneous magnetization.

Measurements were made of the relative changes in electrical resistivity in applied longitudinal and transverse magnetic fields (magnetoresistance). The resistivity of every specimen was found to decrease steadily with increasing magnetization until technical saturation was reached. No further resistivity changes were detected in fields greater than those required to achieve technical saturation at any temperature. This behaviour is in marked contrast to that of ferromagnetic metals (see for example Becker and Döring 1939). Magnetoresistance curves at room temperature for nickel ferrite are shown in figure 1.

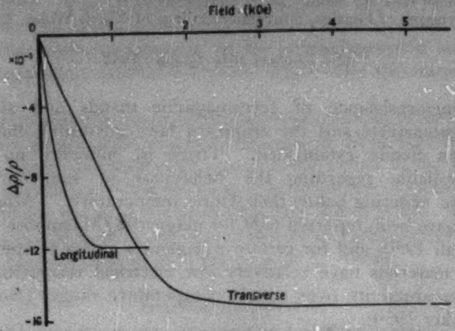


Figure 1.

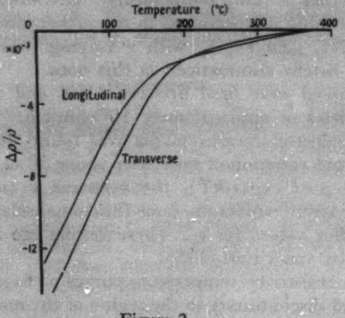


Figure 2.

The relationship between saturation magnetoresistance coefficient and temperature for the same specimen is shown in figure 2. It will be noted that both longitudinal and transverse coefficients fall rapidly with temperature and become immeasurably small above 350°c. Extension of accurate magnetoresistance measurements below 0°c has not proved possible on account of the high specific resistivity of the specimens. Results for ferrites containing zinc are qualitatively similar to those shown in figures 1 and 2; the numerical values of the coefficients at a given temperature, however, were found to decrease steadily with zinc content. It is hoped to present these results in greater detail, together with the methods of measurement, in a later publication.

The results summarized above will now be considered with reference to a theoretical model proposed by Smit (1951) to explain magnetoresistance phenomena. According to Smit the resistivity anisotropy inside a ferromagnetic domain is due to the fact that conduction electrons are subject to a scattering probability which depends upon the angle between their drift velocity and the direction of the magnetization. The calculations made are based upon, and depend wholly on the theory of resistivity of nickel given by Mott (1936). Mott has shown that the resistivity anomaly in nickel resulting from the spontaneous magnetization below the Curie temperature may be accounted for by the enhanced mobility of half of the 4s conduction electrons. It has already been pointed out above that no corresponding anomaly could be found for nickel–zinc ferrites. On the other hand magnetoresistance coefficients of the same order of magnitude as in ferromagnetic metals have been measured. One must conclude therefore that the model proposed by Smit is not adequate to account for the magnetoresistance effects observed in the nickel–zinc ferrite system.

ACKNOWLEDGMENTS

The author wishes to express his thanks to Dr. A. W. Simpson of the Plessey Co., Ltd., for preparing the specimens, to Mr. H. Lord for his assistance with the measurements and to the Council of the Royal Society for an apparatus grant which has enabled him to undertake this investigation.

REFERENCES

BECKER, R., and DÖRING, W., 1939, *Ferromagnetismus* (Berlin : Springer), Chap. 4.
DOMENICALI, C. A., 1950, *Phys. Rev.*, **78**, 458.
GUILLAUD, C., 1950, *J. Phys. Radium*, **12**, 239.
MASUMOTO, H., and SHIRAKAWA, Y., 1941, *Phys. Rev.*, **60**, 835.
MOTT, N. F., 1936, *Proc. Roy. Soc.* A, **153**, 699.
SMIT, J., 1951, *Physica*, **16**, 612.
SMITH, D. O., 1952, *Lab. Insul. Research, M.I.T., Progr. Rep.*, **11**, 53.
VAN SANTEN, J. H., and JONKER, G. H., 1950, *Physical*, **16**, 599.
VAN UITERT, L. G., 1955, *J. Chem. Phys.*, **23**, 1883; 1956, *Ibid.*, **24**, 306.
VOLGER, J., 1954, *Physica*, **20**, 49.

An Explanation of an Anomalous Form of Sub-boundary in Aluminium

BY D. E. BRADLEY AND R. PHILLIPS
Research Laboratory, Associated Electrical Industries Limited, Aldermaston, Berks.

Communicated by G. A. Geach; *MS. received 7th February* 1957

HIRSCH, HORNE AND WHELAN (1956) have reported the direct observations of dislocations in thin foils of aluminium by transmission electron microscopy. Stresses due to the localized high current of the beam in their electron microscope caused the dislocations to move. Similar results have now been obtained in this Laboratory using an experimental prototype Metropolitan-Vickers E.M.6 electron microscope. Apparently anomalous sub-boundaries have been observed which can be explained by the discovery that all dislocations may not be seen with the same contrast.

A specimen of high purity aluminium, beaten down to a thickness of 0·5 micron was annealed at 350°c and etched in one per cent hydrofluoric acid in water to a thickness between 500 and 1000 Å. Portions of the film mounted on copper grids were examined at an instrumental magnification of 55 000, in the electron microscope, operating at 70 kv.

In many cases it was possible to resolve or deduce the arrangement of the dislocations in the sub-boundaries of the aluminium. Almost all the dislocations appeared to be initially in the sub-boundaries, but after a given area had been heated by the beam for approximately a minute dislocations were seen to move out of the boundaries, leaving tracks which, after a few seconds, gradually faded, as described by Hirsch. Multiple glide and cross-slip were frequently observed.

Figure 1 (Plate) illustrates a typical field before the break-up of the boundaries had commenced. A regular, hexagonal network of dislocations is shown at A; at B there is a sub-boundary junction consisting of dislocation nodes; the boundary at C is probably of tilt about two perpendicular axes, that is, it consists of two perpendicular arrays of parallel edge dislocations.

Some dislocations in the network at A, figure 1, are seen with less contrast than others. Similarly, of the three sub-boundaries meeting at A in figure 2 (Plate), Z gives rise to much less contrast than X or Y, although the surface energy and therefore the dislocation content of all three must be similar (Cottrell 1953). Thus it appears that certain types of dislocation, or certain orientations of dislocation with respect to the electron beam, give less contrast than others. This, it is suggested, is the reason for the array seen in the area ABC in figure 3 (Plate). It is well known that such an array of dislocations would not be stable in aluminium. If elements of dislocation line perpendicular to BC and in the plane of the array were not giving much contrast, the dislocation arrangement would be the theoretically expected hexagonal network. The dislocation configuration which probably gave rise to the contrast effects of the area ABC in figure 3 is shown in figure 4, which was obtained by tracing the photograph and drawing in lines perpendicular to BC.

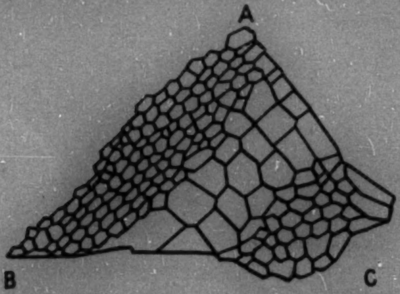


Figure 4. A suggested network to give the contrast effects of figure 3 (×100 000).

It is suggested that the contrast effects at A in figure 5 (Plate) may have a similar interpretation to the region ABC in figure 3. Again the actual dislocation arrangement is probably a hexagonal network in which only certain elements give rise to contrast in the picture. The elements of dark contrast in the region A are arranged in lines which make an angle of 120° with each other, as would be expected on this hypothesis.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge our indebtedness to Dr. Hirsch for giving us help in studying his technique and for supplying the beaten aluminium foil. The authors wish to thank Dr. T. E. Allibone, Director of this Laboratory, for permission to publish this note.

REFERENCES

COTTRELL, A. B., 1953, *Dislocations and Plastic Flow in Crystals*, p. 96, equation 3.23 (Oxford : Clarendon Press).

HIRSCH, P. B., HORNE, R. W., and WHELAN, M. J., 1956, *Phil. Mag.*, **1**, 677.

LETTERS TO THE EDITOR

Would a Classical Gas of Non-attracting Rigid Spheres show a Phase Transition ?

The calculations by Zwanzig (1956) of the first five virial coefficients of gases of rigid oriented squares and cubes have been extended to four dimensions. In the table these results are compared with the consequences of approximations of the 'ring' and 'netted-ring' approximations discussed by Riddell and Uhlenbeck (1953) and Rushbrooke and Scoins (1953). These approximations are based on the neglect, in the Mayer and Mayer (1940) series, of all irreducible cluster integrals except those of two simple topological types. According to Rushbrooke and Scoins (1953), the 'netted ring' approximation is practically equivalent to the best available theories based on the calculation of a distribution function and the use of the superposition approximation. (The latter approach led Kirkwood and Monroe (1941) to predict a transition for a gas of non-attracting rigid spheres.) In the table *b* is the linear dimension of the squares, cubes, etc., while the bottom line indicates the radius of convergence of each approximate virial series. This is obtainable from the work of Rushbrooke and Scoins (1953), and was confirmed by a different method (unpublished) of putting this approximation into closed form.

	Lines			Squares		
	Ring	N.R.	Exact	Ring	N.R.	Exact
B_2	b	b	b	$2b^2$	$2b^2$	$2b^2$
	b^2	b^2	b^2	$3b^4$	$3b^4$	$3b^4$
B_3	$-2b^3$	$+1{\cdot}5b^3$	b^3	$-10{\cdot}67b^6$	$+5{\cdot}67b^6$	$+3{\cdot}67b^6$
B_4	$+3{\cdot}67b^4$	$+2b^4$	b^4	$+36{\cdot}7b^8$	$+8{\cdot}47b^8$	$+3{\cdot}77b^8$
B_5						
$1/v_c$	$1/2b$	$1/\sqrt{3}b$	$1/b$	$1/4b^2$	$1/3b^2$	?

Cubes			Tesseracts		
Ring	N.R.	Exact	Ring	N.R.	Exact
$4b^3$	$4b^3$	$4b^3$	$8b^4$	$8b^4$	$8b^4$
$9b^6$	$9b^6$	$9b^6$	$27b^9$	$27b^9$	$27b^9$
$-56{\cdot}9b^8$	$+19{\cdot}33b^8$	$+11{\cdot}33b^8$	$-303{\cdot}4b^{12}$	$+52{\cdot}4b^{12}$	$+20{\cdot}2b^{12}$
$+351{\cdot}7b^{12}$	$+24{\cdot}87b^{12}$	$+3{\cdot}160b^{12}$	$+3370b^{16}$	$+3{\cdot}21b^{16}$	$-99{\cdot}6b^{16}$
$1/8b^3$	$1/\sqrt{27}b^3$	?	$1/16b^4$	$1/9b^4$	?

The following deductions from the data given in the table can be made:

(*a*) Since B_5 is already negative for tesseracts, while B_4 turns out to be negative for the 'five-dimensional gas', it is to be expected, on the reasonable hypothesis that the exact coefficients are not oscillating violently, but are passing through reasonably smooth maxima as we go up the virial series or alter the number of dimensions, that B_6 or B_7 will be negative for a three-dimensional gas.

(*b*) Since the higher virial coefficients turn out to be small residues resulting from the subtraction of much larger individual cluster integrals, the possibility of violent oscillations cannot be altogether ruled out. Also, the fact that the 'netted-ring' approximation is already going wrong for B_5 is not surprising, and was predicted on other grounds by Nijboer and Fieschi (1953). (Riddell and Uhlenbeck's claim (1953) that the 'ring' approximation gives the correct order of magnitude of the higher virial coefficients is due to a numerical slip.)

(*c*) Since the 'netted-ring' approximation is already going wrong for B_5, and since it leads to a spurious singularity even in the one-dimensional gas, it is to be expected that the integral equation approach will be similarly unreliable in its present form.

(*d*) Therefore, there seems to be *no* reason at present for believing in the physical reality of the 'Kirkwood–Monroe transition'. Riddell and Uhlenbeck (1953) have suggested that it may be connected with the solidification of a gas above its critical temperature, but no molecules behave even approximately as rigid spheres at presently attainable pressures, the attractive forces never being negligible.

Atomic Weapons Research Establishment,
Aldermaston, Berks.
H. N. V. TEMPERLEY.

12th November 1956;
in final form 22nd February 1957.

KIRKWOOD, J. G., and MONROE, E., 1941, *J. Chem. Phys.*, **9**, 514.
MAYER, J. E., and MAYER, M. G., 1940, *Statistical Mechanics* (New York : Wiley).
NIJBOER, B. R. A., and FIESCHI, R., 1953, *Physica*, **19**, 545.
RIDDELL, R. J., and UHLENBECK, G. E., 1953, *J. Chem. Phys.* **21**, 2056.
RUSHBROOKE, G. A., and SCOINS, H. I., 1953, *Proc. Roy. Soc.* A, **216**, 203.
ZWANZIG, R. W., 1956, *J. Chem. Phys.*, **24**, 855.

The Ising Model and Ferrimagnetism

Following the recent applications of the method of high temperature power series to ferro- and antiferromagnetism (Rushbrooke and Wood 1955, Domb and Sykes 1956, Brown and Luttinger 1955, Danielian and Stevens 1957) it is a natural extension to apply it to ferrimagnetism. In the course of examining this an interesting result for an Ising model has been found. Consider a body-centred cubic lattice in which the corners of the unit cell are occupied by ions having spin $\frac{1}{2}$ and g-value g_A, constituting the A sub-lattice, and the centres of the unit cells by ions also with spin $\frac{1}{2}$ but with g-value g_B, forming the B sub-lattice. Nearest neighbour spins are coupled together by an Ising type interaction, $J\, s_z^i s_z^j$, where J is positive and thus favours antiparallel ordering of the two sub-lattices. (It will be noticed that every A ion is coupled only to B ions, and vice versa.) If $g_A = g_B$ the substance will be an antiferromagnetic, but if $g_A \neq g_B$ the onset of ordering will give rise to a net moment, as in a ferrimagnetic.

In the presence of a magnetic field H in the z-direction, the Hamiltonian is the sum of three commuting terms: the exchange energy

$$JW = \sum_{\text{neighbours}} J s_z^i s_z^j$$

and the magnetic energies of the two sub-lattices

$$HM_A = -g_A \beta H \sum_A s_z^i; \quad HM_B = -g_B \beta H \sum_B s_z^j.$$

By the standard methods of statistical mechanics (Daniels 1953) it can be shown that the susceptibility χ can be expressed as

$$\chi(\lambda)=\frac{\lambda}{J}\frac{\mathrm{spur}\,(M_A{}^2+2M_AM_B+M_B{}^2)e^{-\lambda W}}{\mathrm{spur}\,e^{-\lambda W}}$$

where $\lambda=J/\boldsymbol{k}T$. Suppose now that the axes of quantization for all the B ions are reversed, by suitable rotations. This is a purely mathematical device which corresponds to relabelling the states of the ions on the B sub-lattice. No physical change results and the value of the susceptibility cannot be affected. However, the expression for $\chi(\lambda)$ is altered, because both M_B and W are changed in sign. Thus:

$$\chi(\lambda)=\frac{\lambda}{J}\frac{\mathrm{spur}\,(M_A{}^2-2M_AM_B+M_B{}^2)e^{\lambda W}}{\mathrm{spur}\,e^{\lambda W}}.$$

Comparing this with

$$\chi(-\lambda)=-\frac{\lambda}{J}\frac{\mathrm{spur}\,(M_A{}^2+2M_AM_B+M_B{}^2)e^{\lambda W}}{\mathrm{spur}\,e^{\lambda W}}$$

which is the susceptibility of the same system of ions but with a ferromagnetic interaction, it is seen that

$$\chi(\lambda)+\chi(-\lambda)=-\frac{4\lambda}{J}\frac{\mathrm{spur}\,M_AM_Be^{\lambda W}}{\mathrm{spur}\,e^{\lambda W}}$$

and

$$\chi(\lambda)-\chi(-\lambda)=\frac{2\lambda}{J}\frac{\mathrm{spur}\,(M_A{}^2+M_B{}^2)e^{\lambda W}}{\mathrm{spur}\,e^{\lambda W}}.$$

If the same calculation is performed with $g_A=g_B=g$, and the susceptibility is denoted by $\chi_0(\lambda)$, it follows that

$$\chi(\lambda)=\left(\frac{g_A+g_B}{2g}\right)^2\chi_0(\lambda)-\left(\frac{g_A-g_B}{2g}\right)^2\chi_0(-\lambda).$$

This result has a very simple physical interpretation, for it expresses the susceptibility of a ferrimagnetic as the sum of the susceptibilities of an antiferromagnetic, with a g-value equal to the mean of the sum of g_A and g_B, and a ferromagnetic, with g-value equal to the mean of the difference of g_A and g_B. If, as a first approximation, $\chi_0(\lambda)$ is taken as $C/(T+\theta)$, and the mean difference δg between g_A and g_B is taken to be small,

$$\chi\simeq\frac{C}{T+\theta}+(\delta g)^2\frac{C}{T-\theta}.$$

It is then clear that at high temperatures the susceptibility is similar to that of an antiferromagnetic ($\delta g=0$) but that as the temperature is lowered the susceptibility becomes progressively more dependent on the ferromagnetic contribution, which is tending towards infinity as the Curie temperature is approached, the antiferromagnetic contribution varying much less rapidly with T.

The above result is rigorously true for our model, the essential points in the proof being that every A ion is coupled only to B ions, and vice versa, that M_B and W can be reversed by suitable rotations and that A and B ions have topologically identical surroundings. Where these requirements cannot be met, as for example in the corresponding Heisenberg model, the result is not strictly true. Nevertheless, the usual close similarities between the Ising and Heisenberg models suggests that such a relation may in fact be approximately true.

Communicated by L. F. Bates
K. W. H. Stevens.
Department of Physics,
University of Nottingham.
4th February 1957.

Brown, H. A., and Luttinger, J. M., 1955, *Phys. Rev.*, **100**, 685.
Danielian, A., and Stevens, K. W. H., 1957, *Proc. Phys. Soc.* B, **70**, 326.
Daniels, J. M., 1953, *Proc. Phys. Soc.* A, **66**, 673.
Domb, C., and Sykes, M. F., 1956, *Proc. Phys. Soc.*, B, **69**, 486.
Rushbrooke, G. S., and Wood, P. J., 1955, *Proc. Phys. Soc.* A, **68**, 1161.

REVIEWS OF BOOKS

Astronomical Optics and Related Subjects, by Z. Kopal. Pp. xv+428. (Amsterdam: North-Holland, 1956.) 45 guilders.

Professor Kopal and his staff are to be congratulated on their enterprise in organizing this broadly based international symposium, and credit is due to the editor and publishers for the prompt appearance of the volume and the pleasing typographical appearance of its pages.

It has been claimed that optics and astronomy were the two physical sciences which in the last century made the greatest imaginative advances over the mechanistic physics of the day. The subsequent demonstrations of the atomicity of matter and of energy drew scientific imagination towards atomic and nuclear studies, so that there was a tendency in optics and astronomy towards the refinement of what had already been begun, rather than development along essentially new lines. Consequently, ideas which were commonplace in physics as a whole, or in engineering, could become ignored or even denied in optics and astronomy; no neo-Michelson has written '*Photons and their Uses*'.

Be that as it may, and certainly a number of exceptions come to mind, the current resurgence in interest has thrown these two subjects open to the application of many ideas developed in neighbouring disciplines. If the situation is a fruitful one, it is also one which can cause confusion. Inevitably, several workers may be simultaneously developing and propagating an idea, often in different aspects and at different levels. The Manchester Symposium (which, contrary to legend, was almost totally dry) provided a timely and valuable opportunity to exchange points of view and to relate individual work to overall progress in the field. In several cases, discussions begun at the 1954 Florence Meeting were continued, and it is pleasant to notice how often the Proceedings show that authors have reached mutual agreement on matters which were in dispute at the Symposium.

The book is divided into seven sections, and it is possible here to give only an outline of the contents, which with hybrid vigour spill over the formal classification.

Light is the essential physical tool both of optics and of astronomy. It is pointed out that its wave properties limit the number of significant degrees of freedom of an optical image, and its particle properties limit the accuracy with which the independent coefficients can be measured. Thus both the amount and the nature of the information passed by an optical system are fundamentally limited, but within these limits one desirable quality can sometimes be traded for another.

Modern high-grade optical systems for photography commonly have errors amounting to several wavelengths, and their aberrations cannot be discussed satisfactorily either by ray-theory or by the simple methods ordinarily used to assess the harm done by small figuring errors. In the Huygens approximation, the complex amplitudes in the image and aperture planes are essentially Fourier transforms of each other. Fourier transforms have an unjustified reputation for difficulty; their convergence properties may be subtle, but their formalism is simple. Their usefulness extends also to ray-theoretic investigations, since the effect of aberrations is represented by convolution of the image intensity, but by the simpler operation of multiplication of its transform. Accordingly, it is becoming common in optics (as has long been customary in electronics) to regard a sinusoidal distribution rather than a ' point source ' as the element into which object and image distributions can be most usefully analysed. The change is associated with the use of sine-wave response factors to characterize optical systems, and with the use of sinusoidal test targets or interferometric methods to measure these factors. These Fourier methods underlie the discussion of diffraction images and of apodization.

In addition to determining the form of the image, it is necessary for design purposes to abstract from it a measure of image quality. This subject has in the past been somewhat bedevilled by arbitrary choices and by attempts to make a single figure of merit cover all applications, but from the work reported it appears that a figure of merit can now be defined to suit any properly formulated requirement. The need is the more urgent now that propagation through an optical system can be computed automatically on high-speed digital machines, so that if an objective measure of performance is available the machine can seek the most favourable design parameters.

Theoretical developments are reported in the description of the propagation of light solely in terms of observables, with special reference to partial coherence; and in the investigation of situations encountered in wide angle systems and in microwaves where the ordinary Huygens approximation is inadequate.

Considerable attention is paid to the theory, production and application of aspheric surfaces.

In addition to the conceptual and computational facilities referred to, the gifts of electronics to optics include photoelectric methods and the promise (already partly fulfilled) of image devices of greater sensitivity than the photographic plate. It is perhaps a little disappointing that few new data are given on the quantum efficiencies and performance of these devices, and also that only one paper reports a measurement on the noise level of the photographic plate, which is after all still the major light-detector in astronomy.

Many astronomical observations are limited by ' seeing ', the irregular refraction of light in the earth's atmosphere. The systematic study of its effects, which has been somewhat neglected during this century, is taken up in several contributions.

Among new devices for astronomy, improvements in interferometry, in multilayer filters, and in dye filters, are reported; and the perennial problem of matching a round star image to an elongated spectrograph slit receives attention. A new application of the objective-prism method promises to be of the greatest importance in the measurement of radial velocities.

One's chief regret is that the price of the book means that it will be taken up mainly by libraries, and that few individuals (other than reviewers) will have the good fortune to possess a copy of their own. The high standard of production could obviously not have been paid for at a lower price in this relatively small and inflexible market, but one longs, perhaps vainly, for a return to the conditions where a book of this kind could be in the hands of every student of the subject.

P. F.

Physics in my Generation. A selection of papers by MAX BORN, F.R.S., N.L. Pp. viii + 232. (London and New York: Pergamon Press, 1956.) 40*s*.

In 1921, classical physics, enlarged and enriched by the relativity theory of Einstein, seemed to promise an objective picture of the physical universe in which the succession of events was governed by rigid equations of motion and in which apparently irrational features such as Planck's constant would ultimately appear as logical consequences of the fundamental equations. Less than ten years later the foundations of this orderly structure had been undermined, deterministic laws in natural phenomena had been replaced by statistical averages, and the new theory of quantum mechanics had been developed to acknowledge and describe the indeterminacy of atomic processes. Professor Max Born is one of the founders of quantum mechanics; in his early years as a student he was closely in touch with Einstein and Minkowski, and later he numbered among his pupils in Göttingen both Heisenberg and Jordan. His work and experience embrace the period of transition between the classical and quantum manner of thought and in common with many great theoretical physicists he has been much concerned with the philosophical implications of his methods. In the present collection of essays and lectures, nearly all of which have been published elsewhere but which are brought together in this volume for the first time, Professor Born surveys the development of physical thinking over a period of some thirty years or more. The articles are, as might be expected from their various purposes, of widely different length and standard of difficulty, and there is a good deal of repetition, but they are beautifully written and easily readable. Some, notably those on *Cause, Purpose and Economy in Natural Laws* and *Astronomical Recollections*, deviate somewhat from the theme of the majority, and in some of the later lectures there are remarks on the responsibility of scientists for nuclear warfare to which not all physicists will subscribe. The excellence of this book is, however, in the accounts which it gives of the work of the great German physicists, of the difficulty which Einstein and Planck found in accepting the fundamental indeterminacy of nature described by Heisenberg's uncertainty principle, of the reluctance of Schrödinger to accept the reality of anything but wave functions, of the place of Lorentz and Poincaré in the history of relativity, of the unification of alternative descriptions of matter made by Bohr in his principle of complementarity. The physicist learns at first hand how Born brought matrix algebra to Heisenberg's theory of observables, how he was led to identify the square of a wave function with a particle density and how the method of approximation which bears his name became a powerful tool of the new quantum mechanics. All this is of course familiar to many and is to be found tidily presented in numerous textbooks, but most of these fail to convey the excitement of discovery and the spirit of adventure which animated the great men whose names move through Professor Born's pages. Those for whom quantum mechanics is a human achievement as well as a successful technique will find both pleasure and profit in reading and re-reading this book.

W. E. BURCHAM.

The Structure of Turbulent Shear Flow, by A. A. TOWNSEND. Pp. xii + 315. (Cambridge: University Press, 1956.) 40*s*.

'Turbulent' fluid flow, that is, flow which owing to instability has become so irregular that it can be studied only by statistical means, is one of the hardest branches of statistical physics. During the seventy-odd years since Reynolds initiated its study, some sensible things have been said about its influence on the mean flow, and about the character of the 'small eddies' into which the larger eddies break up by a sort of cascade process, and which are responsible for the ultimate viscous dissipation of the turbulent energy. Now at last, for the first time, a treatise appears in which the author, Dr. Alan Townsend of the Cavendish Laboratory, Cambridge, by means of an astonishing combination of experimental ingenuity and fertility of theoretical ideas, has succeeded in giving a reasonably convicing account of the character of the larger eddies which contain the bulk of the turbulent energy, and an explanation on fluid-mechanical principles of their main properties. This is a very original and at the same time very comprehensive book, which, while drawing liberally on the results of the notable aerodynamic research schools at Johns Hopkins University and the California Institute of Technology, and on the ideas of notable past workers such as Prandtl and Taylor, brings to light nevertheless a wealth of new principles and experimental results which, although tentatively expressed in some earlier papers of the author, find here their first coherent exposition.

Dr. Townsend does not write for those whose only interest in the subject is in the technological know-how which they can derive from it, but anyone who wishes to gain real physical insight into turbulent flow processes should make a careful study of his book. M. J. LIGHTHILL.

Wool Research, Vol. 2: *Physical Properties of Wool Fibres and Fabrics*. Pp. 234. (Published by Wool Industries Research Association, Torridon, Headingley, Leeds 6.)

Wool Research was originally planned to make available in book form the many contributions from the Wool Industries Research Association. This second volume, however, is more than a re-publication of earlier papers relating to the physical properties of wool fibres and fabrics, for F. L. Warburton of the Physics Department of the Wool Industries Research Association has rewritten the majority so as to give a concise and up-to-date account of this large field of work.

The scope can be assessed from the following headings of the nine chapters: Relative humidity, regain and swelling; Absorption of water by textile materials; Processes involving absorption and desorption of water; The warmth of clothing; The water repellency of fabrics; Elastic and plastic properties; The electrical properties of wool; Friction; Colour and optical properties of fabrics. There are 204 references over half of which relate to papers from the W.I.R.A.

This volume certainly serves to show the extent to which the work of a Research Associaton has furthered knowledge of the properties and applications of one particular product. It is, however, much more than an account of certain results and conclusions, and succeeds in presenting to the user of wool, and indeed of fibres in general, a clear indication of how the processing and fabricating is dependent on the properties of the fibre and on changes in the environmental conditions. Relative humidity stands out as a factor that is often of major importance.

The subject matter is well presented and there appear to be but few errors. Small slips on pages 41 and 72 escaped the proof-reader.

This volume can be recommended, not only to the textile technologist for whom it is primarily intended, but to all who are interested in clothing and in the important bearing that some sections of classical physics has on this and related subjects.

The seeker for unusual information will be pleased to learn that 1 clo equals 1·55 tog. R. W. POWELL.

CONTENTS OF SECTION A

PAGE

Dr. J. C. Polkinghorne, Dr. D. L. Pursey and Dr. H. A. Santa Maria. K-meson Decays through a Fermi Interaction 321

Prof. S. N. Ghoshal and Mr. B. B. Baliga. Elastic Scattering of 32 mev Protons from Gold 324

Dr. G. W. Greenlees, Mr. B. C. Haywood, Miss L. G. Kuo and Mr. M. Petravic. The Elastic and Inelastic Scattering of Protons by Mg . . 331

Mr. J. Randles. Incoherent Scattering of Gamma Rays from the Bound Electrons of Heavy Elements 337

Dr. K. Takayanagi. Rotational Transition of Hydrogen Molecule by Collision. 348

Mr. K. R. Allen, Mr. E. A. Finlay, Mr. M. Lipsicas, Mr. D. Major and Mr. K. Phillips. The Scattering of Fast Charged Particles: II—On the Single Scattering of 9·8 mev Positrons in Xenon 355

Dr. G. E. Brown. Polarization of gev Protons Scattered by Complex Nuclei . 361

Mr. R. J. Huck. The Effectiveness of Variational Methods for Inelastic Scattering Problems 369

Prof. R. E. Peierls and Dr. J. Yoccoz. The Collective Model of Nuclear Motion 381

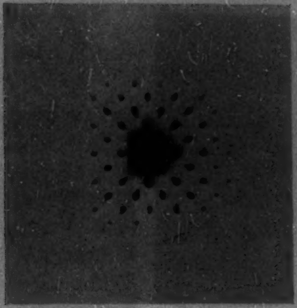
Dr. J. Yoccoz. On the Moments of Inertia of Nuclei 388

Research Notes:

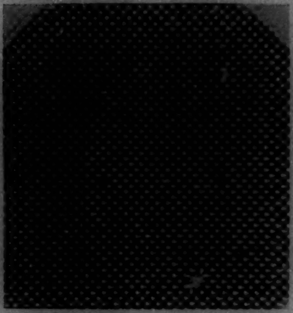
Mr. B. H. Bransden and Mr. J. S. C. McKee. A Variational Calculation of the Cross Section for the 1s–2s Excitation of Hydrogen by Electron Impact 398

Reviews of Books 401

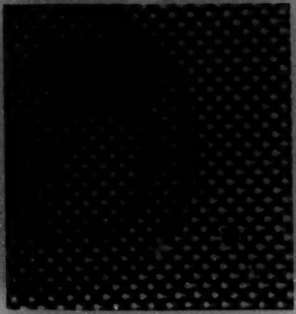
Contents of Section B 404



(*a*)

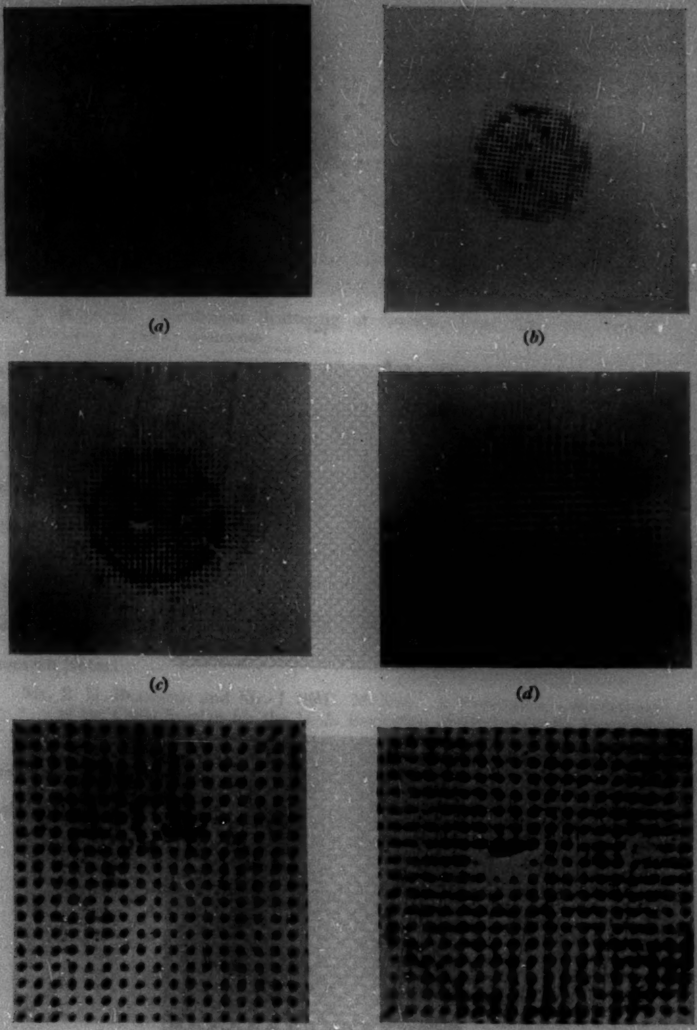


(*b*)



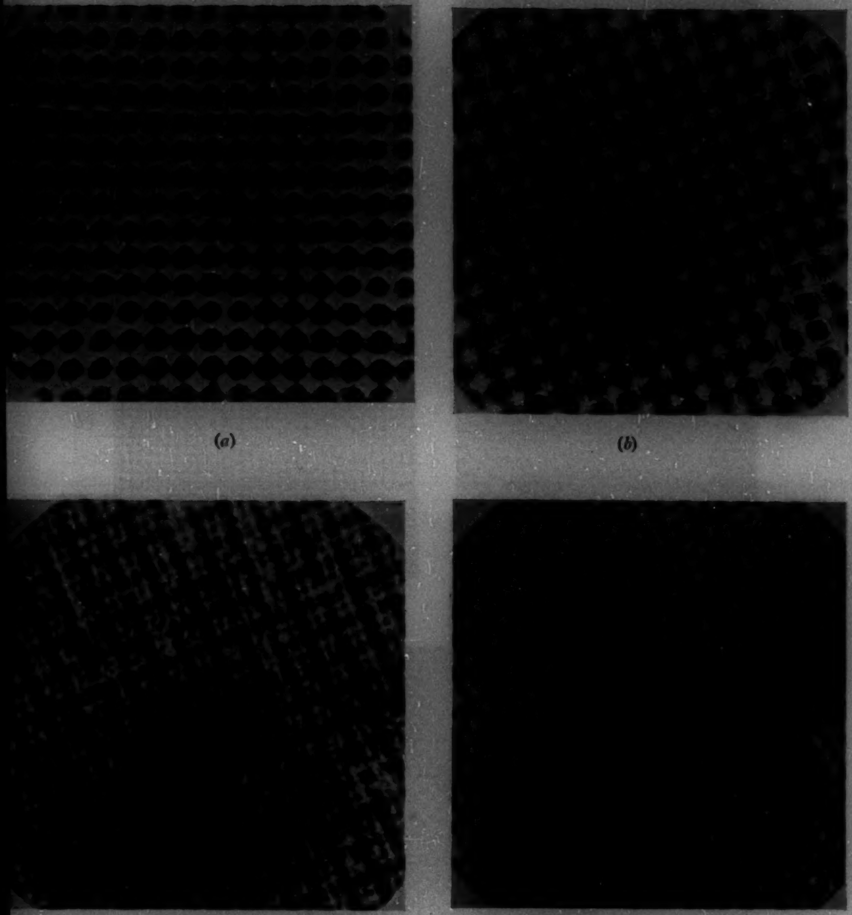
(*c*)

Figure 3. (*a*) Diffraction pattern obtained from the grating of figure 2; (*b*) Magnified image of the grating; (*c*) first Fourier image of the same grating. There is no trace of peak C of figure 2 (*b*).



(*a*) (*b*) (*c*) (*d*) (*e*) (*f*)

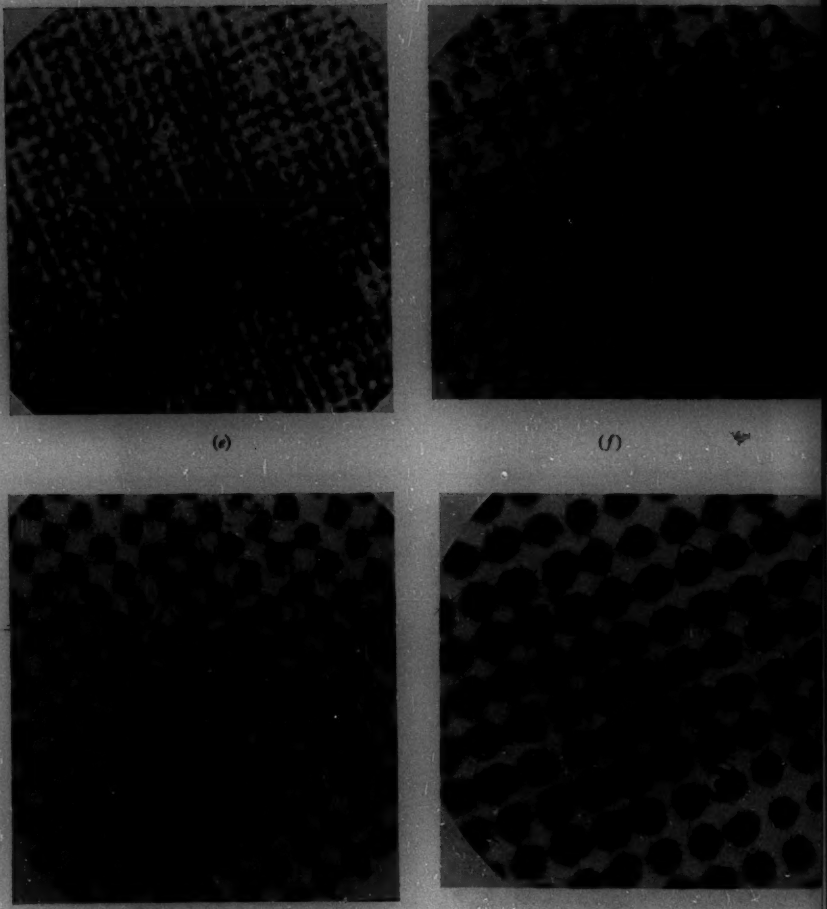
Figure 7. (*a*) $I(-\infty)$, the diffraction pattern; (*b*) $I(-\frac{1}{2})$, the first virtual Fourier image; (*c*) $I(0)$, the object; (*d*) $I(1)$; (*e*) $I(2)$; (*f*) Normal image of the same region at the magnification of $I(2)$.



(*a*) (*b*)

(*c*) (*d*)

Figure 2 (*a*)–(*d*). Magnified image of periodic object. (*a*), *I* (0); (*b*), *I* (0·075); (*c*), *I* (0·142); (*d*), *I* (0·204). See also Plate II.



(*e*) (*f*)

(*g*) (*h*)

Figure 2 (*e*)–(*h*). Magnified range of periodic object. (*e*), I (0·258); (*f*), I (0·310); (*g*), I (0·363); (*h*), I (0·538).

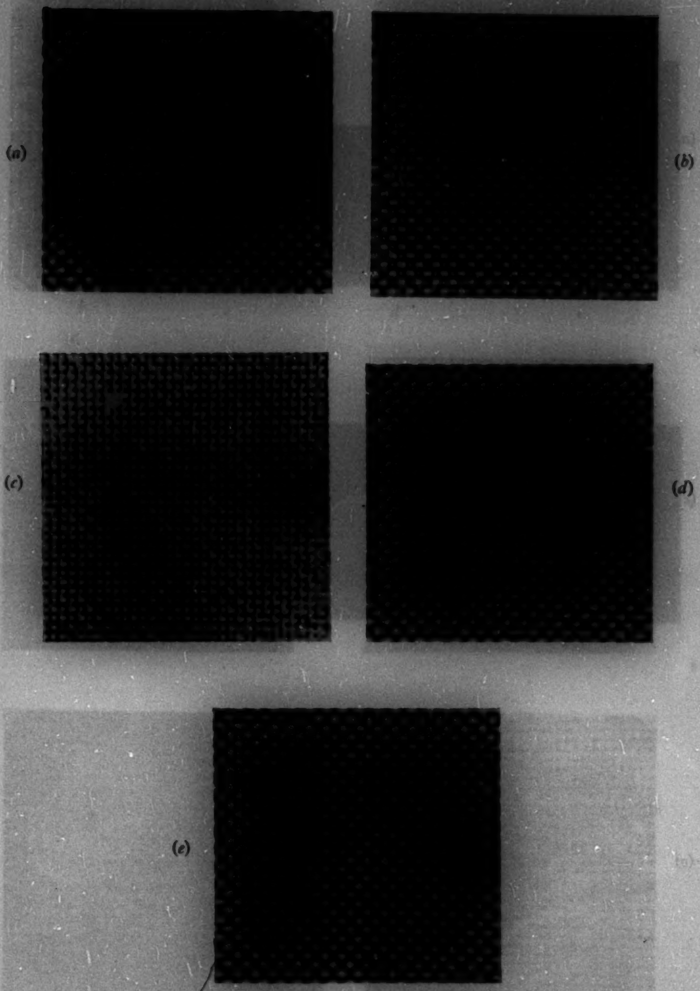


Figure 1. The $\nu=\frac{1}{2}$ Fourier images of an optical grating with $35{\cdot}3\,\mu$ periodicity, obtained with two small sources separated, in a ' horizontal ' direction, by (*a*) $17\,\mu$, (*b*) $46\,\mu$, (*c*) $75\,\mu$, (*d*) $110\,\mu$, (*e*) $143\,\mu$.

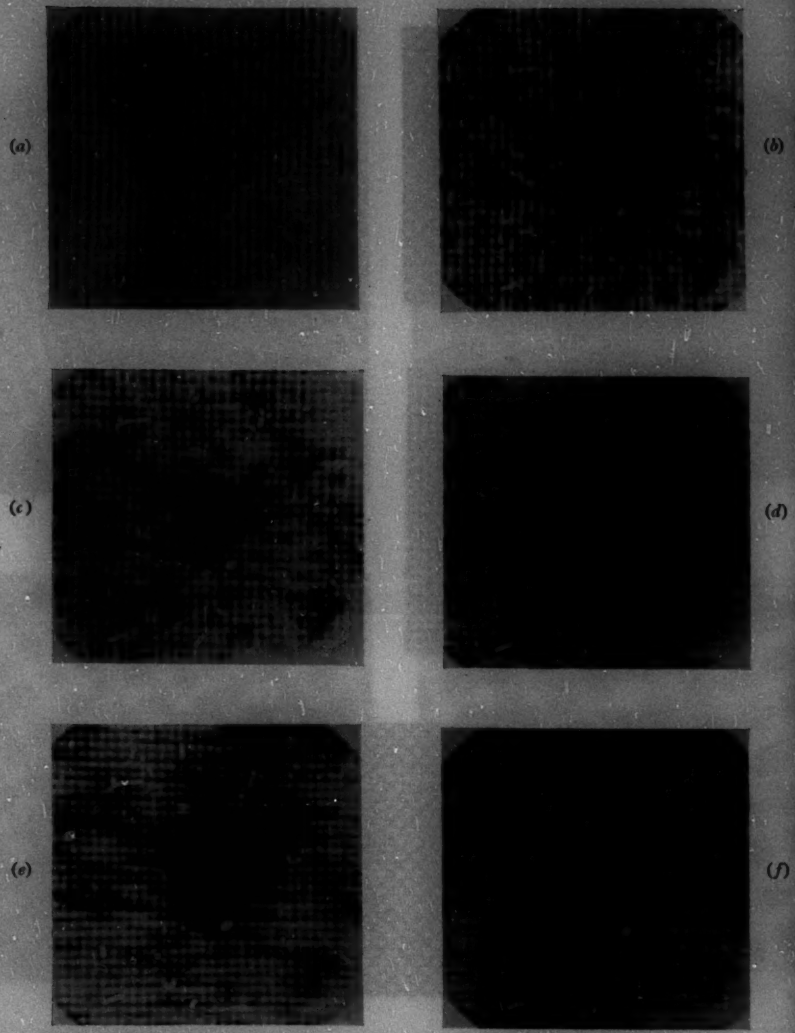


Figure 3. The $\nu=2\frac{1}{2}$ Fourier images of maximum contrast, and the corresponding diffraction patterns, for a grating of $14{\cdot}67\,\mu$ periodicity and a rectangular source of variable length. The photographs are taken with the variable source dimensions (*a*) $13\,\mu$, (*b*) $40\,\mu$, (*c*) $64\,\mu$, (*d*) $93\,\mu$, (*e*) $115\,\mu$, (*f*) $136\,\mu$.

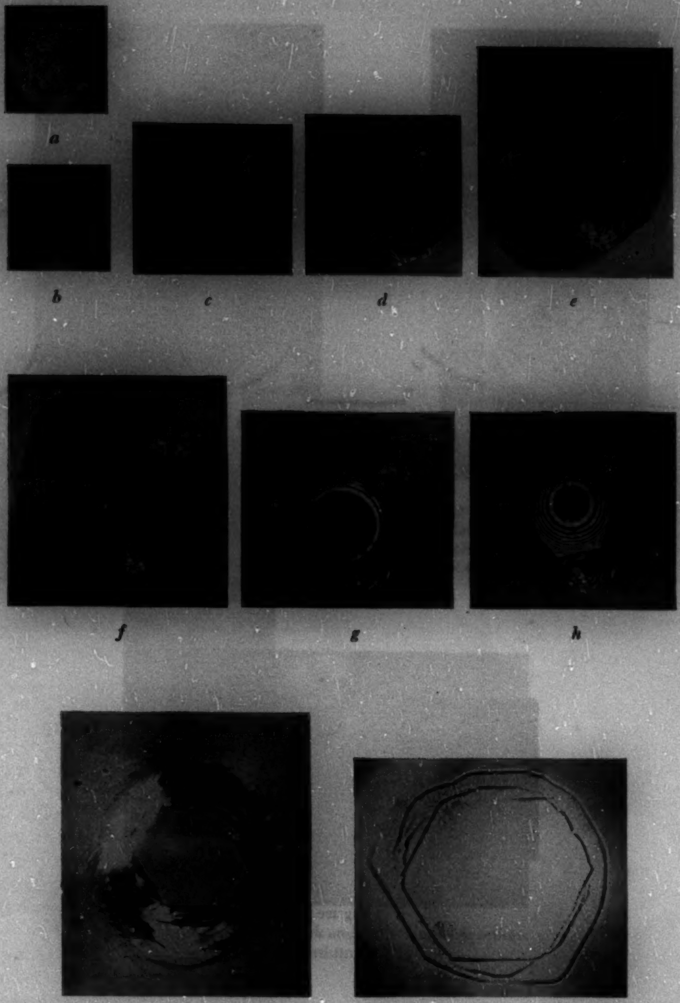
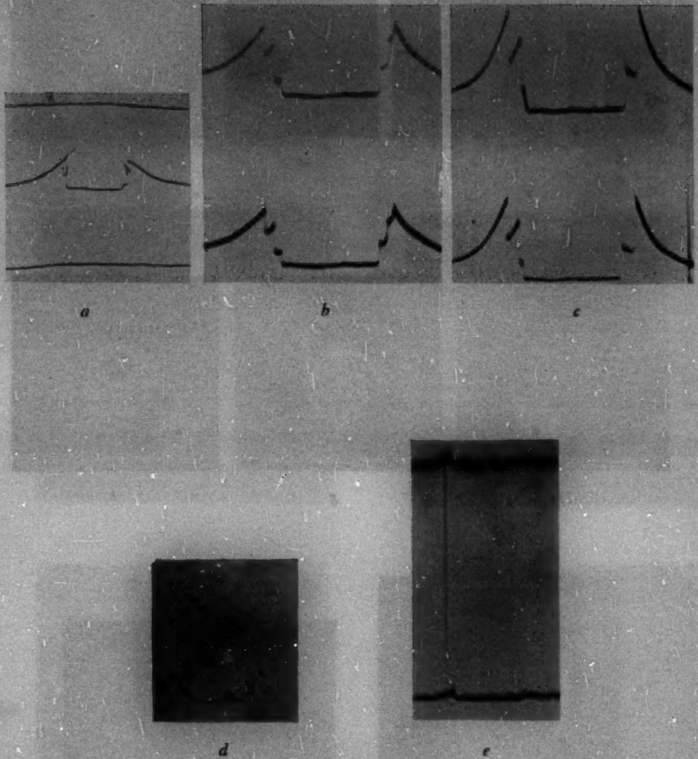

a
b
c
d
e
f
g
h
i
j

Figure 1.



a *b* *c*

d *e*

Figure 3.

Figure 1. Hexagonal and other networks in aluminium (×45 000).

Figure 2. Sub-boundaries of similar dislocation content but differing contrast in aluminium (×70 000).

Figure 3. The contrast effects due to a network of dislocations in aluminium (×45 000).

Figure 5. A network and individual dislocations in aluminium (×78 000).

Self-Diffusion in Gold

By S. M. MAKIN, A. H. ROWE and A. D. LeCLAIRE
Atomic Energy Research Establishment, Harwell, Berks.

MS. received 21st January 1957

Abstract. The self-diffusion coefficient D of 99·95% purity gold has been measured in the temperature range 704°c to 1048°c using the sectioning method. D was found to vary with temperature according to the equation

$$D=(0{\cdot}091\pm0{\cdot}001)\exp[-(41\,700\pm300)/RT]\ \mathrm{cm^2\,sec^{-1}}.$$

A detailed comparison of this result is made with those of earlier workers.

§ 1. Introduction

Self-diffusion of gold in the temperature range 704–1048°c has been investigated by the multiple sectioning technique. 99·95% purity, coarsely polycrystalline, gold discs electroplated with ^{198}Au have been used. ^{198}Au emits β and γ radiation with a 2·69 day half-life and its distribution after annealing was found from measurements of the intensity of 0·411 Mev γ-rays emitted from sections cut from the discs.

Reported measurements of the self-diffusion of gold (Sagrubskii 1937, McKay 1938, Gatos and Kurtz 1954) differ considerably. The present work, which is part of a programme of study on the diffusion of impurities in gold and gold–silver alloys was undertaken in the hope of confirming one or other of these measurements. The results obtained, however, differ from them all but are in quite good agreement with those of Okkerse (1956) and of Mead and Birchenall (1956, private communication), announced before the completion of the present work.

§ 2. Experimental Technique

The experiments were made on 20 mm diameter, 10 mm thick discs of 99·95% gold, pre-annealed at 1000°c for six hours in a vacuum of 10^{-5} mm Hg to standardize the grain size to about 1 grain mm^{-2}. For each diffusion experiment one face of one disc was mechanically polished and then cleaned by electrolytic hydrolysis in 5% caustic soda in alcohol. A layer of active gold, ^{198}Au, about 2×10^{-3} mm thick, was then electroplated on to this surface from an active solution of potassium gold cyanide containing gold at a concentration of 10 mg cm^{-3} and specific activity of 2·0 mc g^{-1}. One sample was prepared by deposition of ^{198}Au from the vapour to provide a variation on the plating technique.

Each specimen, active face down on a flat silica plate, was sealed in a silica tube evacuated to 10^{-5} mm Hg and annealed for a measured time in a tube furnace designed to have a temperature gradient of less than 0·2 deg cm^{-1} over a central 10 cm length. The temperature was controlled to ±0·5°c by using a voltage stabilized power supply and an electronic temperature controller operating from a specially designed platinum resistance thermometer.

This thermometer consisted of a bundle of fine bore silica quills threaded with 0·005 in. platinum wire and sealed parallel to and at one end of a twin bore supporting tube, down which passed heavy platinum leads. This construction gave a steadier control, particularly above 900°c, than thermometers with a ceramic embedded resistance element provided commercially with the electronic controller. This is believed to be due to appreciable capacitative effects in the ceramic at high temperatures (Lomas, Jepson and Rait 1951). The thermometer was permanently installed in the furnace and fitted with locating pins to give automatic location of sample and thermocouple in the constant temperature portion of the furnace.

The temperature was measured with a calibrated 13% rhodium platinum–platinum thermocouple located close to, and at the same level in the furnace as, the specimen. The temperature was recorded continuously during the anneal but accurate measurements were also taken at frequent intervals on a precision potentiometer.

After annealing, the distribution of radioactive material in the specimen was determined by facing off in a lathe a number of slices parallel with the end surface of the specimen and separately counting their activity.

To support the sample for machining, without deforming it, and at the same time to facilitate the aligning of the specimen surface perpendicular to the axis of rotation of the lathe spindle, the inactive face was cemented with 'Araldite' to the flat surface of an adjustable ball joint, which was then held in a collet in the lathe spindle. This device is similar to that described by Gubkin and Dovnar (1953).

It is usual, before slicing, to remove a layer of material from the cylindrical surface of the specimen to avoid the possible effects of surface diffusion. In the present experiments a slightly different procedure was adopted.

Well-annealed gold is very soft and the act of facing produces a peripheral burring of the specimen. This burr grows in extent as successive facing cuts are made so that at any particular stage the burr contains material smeared into it from all previous cuts. Since each cut also removes some metal from the burr the activity of the turnings from any particular cut is not truly representative of the activity of the specimen at the position at which that cut is made, but is a little greater than it should be, and the more so the later the cut.

This effect is avoided if the cylindrical surface material is removed piecemeal as part of each facing cut. The first outer peripheral portions of each cut, annuli 3 mm wide, were therefore rejected and only the turnings from a constant diameter (14 mm) inner cylindrical portion of the specimen were retained for counting.

The thickness of each cut to be taken, except the first, was predetermined by advancing the lathe saddle by the required amount, this being measured on an accurately calibrated band transmission type dial gauge firmly attached to the lathe bed and with its stylus abutting on the end face of the saddle. The thickness of the first cut, which also determines the origin from which all distances are measured, was found from the weight, density and diameter of the cut. Attempts to determine the origin as the position at which the tool appeared first to make contact with the specimen surface were not satisfactory, as the weight of the first cut, taken after advancing the saddle the predetermined distance from this position, frequently differed appreciably from the nearly constant weights of successive cuts.

Cuts were usually 0·02 mm thick. The turnings from the central portion of each cut were collected in separate glass specimen tubes placed in turn under a funnelled hole at the bottom of a Perspex box which enclosed the specimen and cutting tool; a fine air jet was used to sweep the box clean after each cut.

The contents of each tube were weighed, melted into a uniform bead in an induction heater, reweighed, and counted with a scintillation counter. When the weights before and after melting differed, which was very infrequently, the smaller was used in calculating the specific activity (counts g^{-1}). The measured activities were corrected for the counts lost in the recording system, background counts, and the self-absorption of the specimen. Knowing the time at which each specimen had been counted and the half-life of ^{198}Au, the measured activities were converted to activities at a standard reference time.

§ 3. RESULTS

The technique described was used to find the activity distribution in diffusion samples annealed at eleven different temperatures in the range 704–1048°c.

For a specimen consisting initially of a very thin active layer on the end of a long rod the activity at a distance x from the end after annealing at T°K for t sec is given by

$$c = c_0(\pi D t)^{-1/2} \exp(-x^2/4Dt) \qquad \ldots\ldots(1)$$

where c = specific activity at a distance x from the plated surface of the specimen, c_0 = total activity originally deposited, D = diffusion coefficient in cm^2 sec^{-1}. Equation (1) is valid, provided $h^2/12Dt \ll 1$, where h is the thickness of the layer. This condition is satisfied in all the samples used. For each anneal the logarithm of the specific activity was plotted against x^2; a typical set of results is shown in figure 1. The diffusion coefficient was derived from the slope of the calculated line of best fit. The results over the temperature range explored are summarized in table 1.

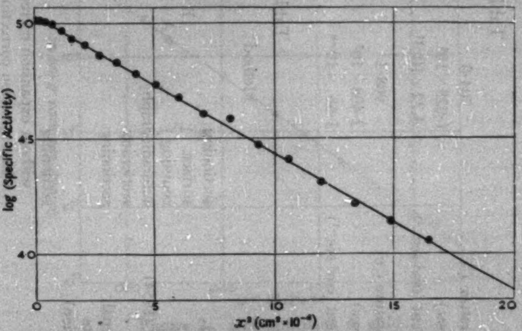


Figure 1. Activity distribution after annealing at 772° c.

The activation energy Q and the frequency factor D_0 are defined by

$$D = D_0 \exp(-Q/RT). \qquad \ldots\ldots(2)$$

Figure 2 shows $\log D$ plotted against $1/T$. Q and D_0 were found from the slope and intercept of the regression line fitting these points.

Table 1. Self-diffusion in Gold

Annealing temperature (°c)	704·0	742·5	772·0	802·0	836·5
Annealing time (sec)	$1\cdot029\times10^{6}$	$9\cdot356\times10^{5}$	$1\cdot033\times10^{6}$	$5\cdot225\times10^{5}$	$5\cdot323\times10^{5}$
Diffusion coefficient ($cm^2\ sec^{-1}$)	4.22×10^{-11}	$9\cdot76\times10^{-11}$	$1\cdot80\times10^{-10}$	$3\cdot01\times10^{-10}$	$5\cdot56\ \times10^{-10}$
Annealing temperature (°c)	908·5	913·5	948·5	1011·0	1048·0
Annealing time (sec)	$2\cdot408\times10^{5}$	$8\cdot474\times10^{4}$	$1\cdot728\times10^{5}$	$7\cdot922\times10^{4}$	$1\cdot762\times10^{5}$
Diffusion coefficient ($cm^2\ sec^{-1}$)	$1\cdot66\ \times10^{-9}$	$1\cdot81\times10^{-9}$	$3\cdot29\times10^{-9}$	$7\cdot26\times10^{-9}$	$1\cdot15\times10^{-8}$

Table 2. Self-diffusion in Gold

Author	Method		D_0 ($cm^2\ sec^{-1}$)	Q (kcal mol^{-1})	D(1000° K) ($cm^2\ sec^{-1}$)	D(1250° K) ($cm^2\ sec^{-1}$)
Sagrubskii (1937)	sectioning		0·16	53·0	$4\cdot14\times10^{-13}$	$8\cdot52\times10^{-11}$
McKay (1938)	surface	B	2·10	50·3	$2\cdot02\times10^{-11}$	$3\cdot23\times10^{-9}$
(recalculated)	counting	A+B	0·050	40·8	$5\cdot97\times10^{-11}$	$3\cdot64\times10^{-9}$
Gatos and Kurtz (1954)	autoradiography		0·265	45·3	$3\cdot32\times10^{-11}$	$3\cdot18\times10^{-9}$
Okkerse (1956)	sectioning		0·031	39·4	$7\cdot58\times10^{-11}$	$4\cdot00\times10^{-9}$
Mead and Birchenall (1956, private communication)	sectioning		0·14	42·9	$5\cdot97\times10^{-11}$	$4\cdot48\times10^{-9}$
Present work	sectioning					
	without expansion correction		0·087	41·65	$6\cdot86\times10^{-11}$	$4\cdot54\times10^{-9}$
	with expansion correction		0·091	41·7	$6\cdot99\times10^{-11}$	$4\cdot65\times10^{-9}$

The distances x in equation (1) should be those obtaining at the annealing temperature, whereas all measurements of distance were made at room temperature. Room temperature distances give an error in D of about 3% at the high end and about 2% at the low end of the temperature range. Table 2 indicates the values of D_0 and Q computed from the exact values of D.

At the higher temperatures, where correspondingly short annealing times were used, the time taken for the specimen to reach the annealing temperature, A, was a significant fraction of the total time spent in the furnace. In the calculation an effective total time of anneal t_e at A°K was used. This is defined by the equation

$$D_A t_e = \int_0^{t_t} D_{T(t)}\,dt$$

where t_t = time between inserting and removing the specimen from the furnace. Thus

$$t_e = \frac{\int_0^{t_t} \exp[-QR/T(t)]\,dt}{\exp[-Q/RA]}.$$

t_e was calculated from this equation using an approximate value of Q. The integral was evaluated graphically from the detailed temperature readings obtained during the heating up. A cooling correction was unnecessary as the specimen cooled very rapidly to temperatures at which diffusion was insignificant.

For the one sample prepared by vapour deposition the diffusion coefficient at 941·0°C was found to be $2{\cdot}73 \times 10^{-9}$ cm^2 sec^{-1}, a value which lies on the line in figure 2. This indicates that the results obtained are not dependent upon the methods used to lay down the active layer.

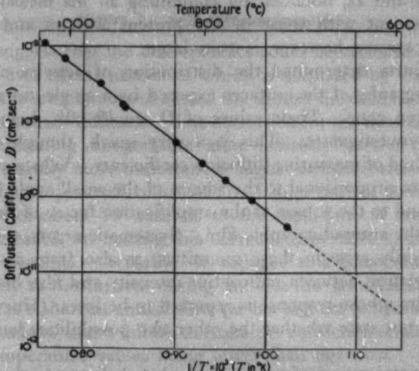


Figure 2. Self-diffusion in gold.

The method of least squares has been used to fit all the results in this paper The variance of the data has been calculated and the standard deviation is quoted as a measure of the errors in D_0 and Q. These values only take random errors into account. Systematic errors have been estimated from the calibration of the dial gauge and on an assumption of a 1°C error in the temperature measurement. They cause less than 1% error in Q and about 20% error in D_0.

§ 4. Discussion

The available data on the self-diffusion coefficient of gold are shown in table 2. The present results and those of Mead and Birchenall agree with one another much more closely than do any other pair of results.

We shall first discuss possible sources of error in the variously reported values of D and of D_0 and Q.

The values of D reported by Sagrubskii are about 60 times smaller than any of the other results. Sagrubskii used samples consisting of a radioactive gold foil plated on both sides with inactive gold. After anneal, successive layers of gold were dissolved off and their activities compared with the activity of a portion cut from the central foil before it was plated. We should expect each dissolved layer to have an activity greater than the one dissolved off before it, but this is by no means a universal feature of the detailed experimental results presented by Sagrubskii in his table 3. Sagrubskii remarks that the diminution in activity of his samples over the period of an experiment, never greater than 30 hours, was not serious. If by this he means that no correction was ever made for decay we have a possible explanation of the behaviour in his table 3 and of his low D values. The behaviour in table 3 could be due to the activities of the later dissolved layers having been reduced by decay to values even below the activities of the earlier layers, while values of D even 50 times smaller than they should be could easily result if the activities of the dissolved layers were compared with that of the central foil measured before plating and annealing.

McKay's values of Q and D_0 are derived from measurements on samples prepared by only one of the two methods he used. The results from the other method were quoted but rejected in the calculation because of their larger scatter. The values of Q and D_0 obtained by combining all his measurements are in fairly good agreement with those of the present authors and of Mead and Birchenall. The scatter, however, is quite large.

Gatos and Kurtz determined the distribution of activity after anneal by contact autoradiography of the surfaces exposed by a single cut made obliquely across the diffusion zone. Their values of D are 40–60% below those of the remaining three investigators. This is a very quick, though not necessarily very precise, method of measuring diffusion coefficients. Values of D determined by this method are proportional to the square of the small angle ($\simeq 3°$) at which the cut is made and to the square of the amplification factor of the densitometer used to monitor the autoradiographic film. Systematic errors, the most difficult to trace, could easily arise in these quantities, as also from a departure from linearity in the relation between radioactive intensity and film blackening. The authors used a film whose response is reported to be linear (Hurwitz 1955), but unfortunately do not state whether the other two possibilities have been looked into.

The three remaining investigators all employed the multiple sectioning and counting technique, which is generally believed to be the most reliable method for self-diffusion measurements. Even where the differences are greatest, all the values of D agree with the authors' to within 14% and are much closer than this over the greater part of the temperature range. This, together with what support might be claimed from McKay's average values, strongly suggests these measurements are close to the real answer. Small systematic differences, leading to different values of Q and D_0, still remain however.

Mead and Birchenall measured D at only four temperatures but their points are disposed about our line much more closely than they are about Okkerse's, as can be seen in figure 3. The only remaining discrepancy is therefore between the present results and those of Okkerse. At the highest temperature employed by Okkerse his value is 14% lower and at the lowest temperature 10% greater than ours. Both sets of experiments relate to measurements at about 10 different temperatures and there is no noticeable difference in the accuracy of fit of the points on the respective plots of $\log D$ against $1/T$, cf. figures 2 and 3.

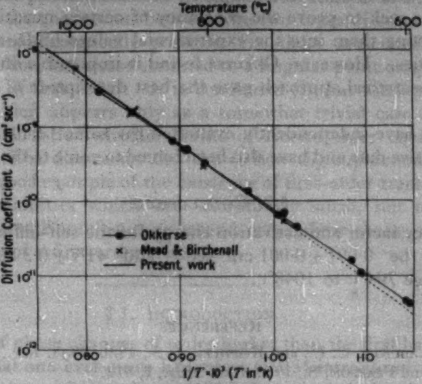


Figure 3. Self-diffusion in gold.

We would suggest that the discrepancies could be accounted for as due to a combination of two effects neither of which is discussed by Okkerse.

The higher the temperature of anneal the shorter the time necessary to achieve a measurable amount of diffusion. The correction for the time taken for a sample to reach the annealing temperature after insertion in the furnace becomes therefore progressively more important at higher temperatures, and if not made will lead to erroneously small values of D. Under normal conditions we found that corrections of about 10^3 seconds to the total time of anneal could arise. D would be increased, on making this correction, by $\simeq 10\%$ in anneals of total duration $\simeq 10^4$ seconds. Actually, we reduced this considerably by boosting the power supply to the furnace during the first few minutes after inserting the sample. Okkerse's shortest anneal time (at his highest temperature) was of this order of magnitude so that a failure to make the correction could largely account for his high temperature values being lower than our own. The effect at the lower temperatures and longer times is negligible, and here Okkerse's values are greater than ours. This could be due to burring, the effects of which, if not dealt with in the manner we have described, and Okkerse did not do so, are to produce an erroneously high value of D at all temperatures. At low temperatures this is the only effect. At high temperatures this combined with the effects of an uncorrected annealing time could be responsible for the observed discrepancies.

The activation energy for self-diffusion of $1{\cdot}81 \pm 0{\cdot}01$ ev in the present work can be compared with the sum of the energy for formation and the activation energy of movement of vacancies found by electrical resistivity measurements on quenched gold wires. Kauffman and Koehler (1955) obtained a total of 1·96 ± 0·06 ev, Bauerle, Klabunde and Koehler (1956) 1·68 ± 0·12 ev, and Bradshaw and Pearson (1956 and private communication) 1·63 ev. The most recent and more reliable values agree quite well with one another but are distinctly less than the self-diffusion activation energy. The same is also true for platinum (Kidson, to be published, Bradshaw and Pearson 1956).

Okkerse (1956) has gone to considerable trouble to compare recent self-diffusion data on face-centred cubic metals with theoretical treatments of diffusion. Such treatments seek to prove the constancy of certain quantities involving D_0 and Q, by inserting them into the experimental values of D_0 and Q, and other physical properties. However, Okkerse found it impossible on these grounds to decide which theoretical approach gave the best description of the experimental results.

The authors have independently evaluated the same 'constants' using more recent and extensive data and have also been forced to come to the same conclusion.

§ 5. Conclusions

The frequency factor and activation energy for the self-diffusion of gold have been found to be $0{\cdot}091 \pm 0{\cdot}001$ cm^2 sec^{-1} and $41{\cdot}7 \pm 0{\cdot}3$ kcal mol^{-1} in the temperature range 704°c to 1048°c.

References

Bauerle, J. E., Klabunde, C. E., and Koehler, J. S., 1956, *Phys. Rev.*, **102**, 1182.
Bradshaw, F. J., and Pearson, S., 1956, *Phil. Mag.*, **1**, 812.
Gatos, H. C., and Kurtz, A. D., 1954, *J. Metals, N.Y.*, **6**, 616.
Gubkin, S. I., and Dovnar, S. A., 1953, *Dokl.Akad. Nauk. SSSR*, **93**, 1025, AEC-tr-2249.
Hurwitz, J. K., 1955, *Rev. Sci. Instrum.*, **26**, 518.
Kauffman, J. W., and Koehler, J. S., 1955, *Phys. Rev.*, **97**, 555.
Lomas, T. W., Jepson, M. D., and Rait, J. R., 1951, *J. Iron St. Inst.*, **168**, 126.
McKay, H. A. C., 1938, *Trans. Faraday Soc.*, **34**, 845.
Okkerse, B., 1956, *Phys. Rev.*, **103**, 1246.
Sagrubskii, A. M., 1937, *Izv. Akad. Nauk. SSSR*, **6**, 903.

Phase Changes of the Second Order

By C. HURST
Jesus College, Oxford

MS. received 7th December 1956

Abstract. The ideal phase transition in superconductors is analysed in such a way as to examine the connection between the two types of second-order phase change which are mathematically possible, viz. those associated with a finite length of transition line, and those confined to a singular point on a transition line which is elsewhere of first order. It is concluded (i) that the latter category of second-order transition appears only as a somewhat trivial case of the former, (ii) that it is very doubtful whether points of second-order equilibrium really exist in isolation. The discussion also shows how the superconducting phenomenon provides a good example of the existence of first-order transitions possessing volume discontinuities which are exceedingly small, but non-zero; this circumstance allows an Ehrenfest-type second-order phase change to be regarded as the limiting case of a first-order transition.

§ 1. Introduction

The subject of phase changes of order higher than the first has proved to be a controversial one ever since Ehrenfest (1933) introduced the concept of second-order transitions. Lype (1946) investigated the thermodynamics of higher-order equilibria by using expansions of the Gibbs function $G(P, T)$ in the form of Taylor series assumed to be valid in the neighbourhood of a point in the (P, T) diagram, and thereby obtained results at variance with those deduced by Ehrenfest. It was pointed out by Bridgman (1946) that these discrepancies were due to a fundamental difference in the viewpoints held by these authors about the physical nature of a second-order transition: Lype's mathematics was appropriate to a second-order equilibrium existing at an isolated point in the (P, T) plane, the transition at every other point in the plane being a first-order one, whereas the treatment of Ehrenfest was based upon the assumption that two phases were in second-order equilibrium along a *finite* length of transition curve. Bridgman expressed the opinion that the state of affairs envisaged by Lype, demanding a second-order transition at a point of intersection of two first-order transition lines, was highly exceptional, and that there was no experimental evidence for its existence. Nevertheless, subsequent workers have continued to discuss the thermodynamic formulae applicable to an isolated second-order point on an otherwise first-order equilibrium line. In particular, Price (1951) has investigated the possibility of treating the phenomena of the vapour–liquid critical point and of the helium lambda-point on this basis, and van der Leeden (1955) has attempted a systematic classification of the various possible types of transition of two-phase systems in which the concept of a high-order isolated point on a lower-order transition line plays a prominent role. Largely as a result of the work of Price and of Zimm (1951), it now seems to be generally agreed that there is no known example in the (P, T) plane of a second-order equilibrium at an isolated point on a first-order line. It is, however, frequently stated (e.g. van der Leeden 1955, Temperley 1956, p. 10) that the superconducting transition does provide an instance, in the variables H and T (where H is the applied magnetic field), of an isolated second-order transition occurring at the zero-field intersection of the two first-order transition curves which correspond to positive and negative values of H. Temperley, indeed, goes so far as to suggest (1956, p. 11) that the existence in practice of this type of second-order transition has now been more clearly established than has the existence of a second-order equilibrium of the original Ehrenfest kind. The purpose of the present paper is to examine the superconducting transition in such a way as to reveal the interrelation of the two postulated varieties of phase changes of the second order; as one of its results, the discussion will show that the above remark of Temperley is certainly invalid.

§ 2. NOMENCLATURE

For the sake of brevity, it will be convenient to follow the nomenclature adopted by Temperley (1956) for classifying changes of phase into thermodynamic types. A first-order transition is said to be of Class A, a second-order one to belong to Class B; when it is illegitimate to represent $G(P, T)$ as a Taylor series owing to the existence of a 'stop-point' on the equilibrium curve, or because second-order derivatives of G become infinite, or for any other reason, then the corresponding transition is referred to as being of Class C. We exclude from the present discussion all phase changes of this last category, an example of which occurs at the critical point of a gas, and also phase changes of order higher than the second. A second-order transition is classified as a Type B1 transition if it pertains to a finite length of the equilibrium curve, as envisaged by Ehrenfest, and as a Type B2 transition if it is confined to a single isolated point on an otherwise first-order equilibrium line.

For purposes of comparison, it should be observed that van der Leeden (1955) uses the notation D(1, 1) to characterize a Class B2 change of phase in the general case when the first-order lines intersecting in the singular point are distinct. The mathematical possibility of having a degenerate case in which these first-order lines become coincident is catered for by introducing the symbolism $\mathrm{D}(\overline{1, 1})$. It turns out that the mathematical relations necessary for such an eventuality are formally identical with those governing a Class B1 transition, but it must be clearly realized that the range of validity of these relations is quite different in the two cases: for a B1 equilibrium, the thermodynamic equations hold good along the length of a line, while for a $\mathrm{D}(\overline{1, 1})$ transition they are true only at an isolated point.

§ 3. THE IDEAL PHASE TRANSITION IN SUPERCONDUCTORS

In investigating the thermodynamics of the superconducting transition, the procedure often adopted is to treat the pressure P and the specific volume V of the metal as remaining strictly constant throughout the analysis. The condition for phase equilibrium is derived by using the 'magnetic Gibbs function' G_m, defined for unit mass by the equations

$$G_m = E - TS + PV - HIV = G - HIV,$$

where I denotes the (uniform) magnetic moment per unit volume, H the (uniform) applied magnetic field, and the notation is otherwise self-explanatory. The function G, and hence the function G_m, is regarded as being dependent only on the variables H and T, so that attention is focused throughout on happenings in one (H, T) diagram. Such a restricted treatment can lead to misconceptions and erroneous conclusions when questions involving the nature of phase transitions are at stake. To obtain a clear picture, it is necessary to consider the complete H–T–P surface over which the normal and the superconducting phases co-exist in equilibrium, rather than just a two-dimensional equilibrium curve isolated from that surface by selecting a plane section of it.

We therefore consider unit mass of a given superconducting metal of such a shape that the demagnetizing coefficient is zero and suppose that (H_c, T, P) and $(H_c + dH_c, T + dT, P + dP)$ are two neighbouring points on the surface which separates the normal from the superconducting phase, and on which the field therefore assumes 'critical' values. In passing from one point to the other the increment of the magnetic Gibbs function is

$$dG_m = V\,dP - S\,dT - VI\,dH_c + \mathrm{O}(2),$$

where the last term comprises all second- and higher-order terms in dP, dT, dH_c. Since the two phases are in equilibrium at every point on the surface the value of dG_m must be the same for each phase. Consequently

$$(V_n - V_s)dP - (S_n - S_s)dT - (V_n I_n - V_s I_s)dH_c = \mathrm{O}(2),$$

where the suffix n refers to the normal phase at a point on the equilibrium surface and the suffix s refers to the superconducting phase at this same point. In the ideal transition under discussion, it is assumed that the magnetic induction is zero throughout the whole volume of the superconducting phase, no account being taken of the possible effects of field penetration which have recently been investigated by Pippard (1956). Putting $I_s = -H_c/4\pi$ and, as usual, $I_n = 0$ for the normal phase, the above equation reduces to

$$(V_n - V_s)dP - (S_n - S_s)dT - (V_s H_c/4\pi)dH_c = \mathrm{O}(2). \qquad \ldots\ldots(1)$$

Since the critical field strength H_c is a function of the two independent variables P and T equation (1) is equivalent to

$$\left\{(V_n - V_s) - \frac{V_s H_c}{4\pi}\left(\frac{\partial H_c}{\partial P}\right)_T\right\}dP - \left\{(S_n - S_s) + \frac{V_s H_c}{4\pi}\left(\frac{\partial H_c}{\partial T}\right)_P\right\}dT = \mathrm{O}(2).$$

The phases will be in equilibrium at the point (H_c, T, P) on the interphase surface if, and only if, the left-hand side of this equation vanishes for all possible independent variations of the small increments dP, dT. If L is the latent heat of the transition, we therefore have

$$\left.\begin{aligned} L \equiv T(S_n - S_s) &= -\frac{TV_s H_c}{4\pi}\left(\frac{\partial H_c}{\partial T}\right)_P \\ \text{and} \qquad V_n - V_s &= \frac{V_s H_c}{4\pi}\left(\frac{\partial H_c}{\partial P}\right)_T \end{aligned}\right\} \qquad \ldots\ldots(2)$$

as *two* conditions which are necessary and sufficient for phase equilibrium at (H_c, T, P). These are standard equations relating to a transition which is usually of Class A but which, in special circumstances, may be of Class B2. The equations are derived, by an alternative method to the one adopted here, by Shoenberg (1952, pp. 59, 74); in his treatment, however, the first equation of the pair is obtained as the result of a thermodynamic argument in which the specific volume V of the metal is regarded as a constant. An immediate consequence of equations (2) is the usual Clausius–Clapeyron equation

$$\left(\frac{\partial P}{\partial T}\right)_{H_c} = \frac{L}{T(V_n - V_s)},$$

relating here to increments of pressure and temperature which are so adjusted as to keep the imposed magnetic field constant and critical.

The transition becomes one of second order at the single point (H_c, T, P) whenever the values of L and of $V_n - V_s$ are zero simultaneously at that point. As is immediately obvious from equations (2), taken in conjunction with the experimental fact that both $(\partial H_c/\partial T)_P$ and $(\partial H_c/\partial P)_T$ are in all cases finite, the necessary and sufficient condition for the desired state of affairs is $H_c = 0$.

Granted that the equilibrium is not of Class C or of order higher than the second, we have now established the fact that it will be of Class B2 if, and only if, $H_c = 0$ at the point in question. If any curve whatsoever be drawn on the interphase surface so as to cross just once the plane of zero field, then the equilibrium at every point on this curve will be of Class A, save for the single point on the curve at which $H_c = 0$: at this isolated point on the curve the transition is one of Class B2.

Let us now fix attention on the equilibrium line on the interphase surface which lies entirely in the plane $H = 0$, i.e. the curve which defines the variation with pressure of the transition temperature T_c of the superconductor, and which has been realized in practice in experiments such as those of Chester and Jones (1953) on tin over the pressure range from 1 to 18000 atmospheres. As has been demonstrated, every single point on this line corresponds to a Class B2 transition. But clearly, as a mere consequence of this, the whole line must itself represent a second-order transition of the true Ehrenfest type. Any assertion that the superconducting phenomenon provides an example of a Class B2 transition must automatically imply the truth of the assertion that this phenomenon also provides an instance of a Class B1 transition. Temperley's statement, already quoted, to the effect that the experimental evidence in favour of the actual existence of a Class B1 equilibrium is less convincing than is that supporting the existence of Class B2 equilibria is seen to be untenable; and, in addition, it appears that the line-form of second-order transition contemplated by Ehrenfest is much more fundamental in character than is the other.

Finally, let us consider the (P, T) equilibrium line obtained by taking the section of the interphase surface by the plane $H_c = h$, where h is a positive constant. This is a Class A transition line for all values of h. However, because H_c has finite first derivatives, it is clear from equations (2) that, by taking h sufficiently small, the magnitudes of the latent heat and the volume discontinuity at transition can be made less than any pre-assigned positive number, no matter how small it may be. Thus we achieve an example of the class of 'small Δv' equilibria, the concept of which was introduced (Hurst 1955) in the course of an investigation of the relationship between Class A and Class B1 transitions. The operation of achieving a second-order transition of the Ehrenfest type as the limiting case of a set of small Δv transitions may, in the present instance, be clearly visualized by considering the sequence of (P, T) curves which is obtained as the plane of section is gradually depressed until it eventually coincides with the plane of zero field.

§ 4. CONCLUSIONS

It has already been remarked that no instance of a Class B2 transition involving only the basic variables P and T has yet been discovered. The above considerations have shown that one particular system needing the introduction of a third independent thermodynamic variable for its complete description displays a Class B2 transition only as a rather trivial aspect of a Class B1 transition, this latter being the locus, in the (P, T) plane for which $H=0$, of all possible Class B2 transitions which could be obtained by restricting attention in succession to the various members of a suitably chosen set of plane sections of the surface over which the two phases co-exist in equilibrium. It is tempting to conjecture that this state of affairs is quite general, and that, in all instances, the observation of a Class B2 transition will arise merely from an inflexibility of the experimental conditions pertaining to that observation, this inflexibility resulting in the excision of just one point from a second-order line. If it indeed be the case that in nature points of second-order phase equilibrium do not exist in isolation, but rather always aggregate to form continuous second-order equilibrium lines, then it would appear unnecessary to subdivide the class of second-order transitions into Types 1 and 2. On this point, it is relevant to remark that, so far as is known, first-order transitions do not happen at isolated points in the (P, T) diagram, but always along continuous curves. It seems not unreasonable that the same should be true of second-order transitions. It is suggested that, until such time as there is definite experimental evidence for the existence of a second-order transition confined to a point in space of the appropriate number of dimensions (this being the number of pairs of conjugated variables required for a full thermodynamic treatment), it is undesirable to cater for such a type of transition in the scheme of classification.

The above considerations are in no way vitiated by the analysis of Pippard (1956) on the effect of field penetration on the superconducting phase transition. Since Pippard is obliged to consider a specific model of a superconductor in order to take account of the variation of penetration depth with field his conclusions cannot be regarded as infallible. But his treatment, which is confined throughout to the (H_c, T) diagram, does suggest that the usual statement that there exists a Class B2 transition at the zero-field point on the equilibrium curve in this diagram is probably incorrect. Pippard finds instead that the transition lines for positive and negative magnetic field are continuous in all their derivatives at this point; but that, at a temperature slightly less than T_c, each transition line contains a critical point marking a change from a transition of the second order to one of the first order. In other words, the model implies that, even in the (H_c, T) diagram itself, there is a short line-element along which the transition is of Ehrenfest's type. Thus, Pippard's work tends to confirm the thesis that isolated points of second-order equilibrium do not occur in practice.

It appears that, in all types of phase change, lines play a more important and fundamental role than do points. This being so, it is not surprising that considerable discretion must be exercised when adopting the device of expanding a function as a Taylor series in the neighbourhood of a point as the basis of a systematic study of the various categories of phase transition: uncritical use of the method is liable to lead to results which, though mathematically sound, are of little or no physical interest.

REFERENCES

BRIDGMAN, P. W., 1946, *Phys. Rev.*, **70**, 425.
CHESTER, P. F., and JONES, G. O., 1953, *Phil. Mag.*, **44**, 1281.
EHRENFEST, P., 1933, *Comm. Kamerlingh Onnes Lab. Leiden, Suppl.* 75 b.
HURST, C., 1955, *Proc. Phys. Soc.* B, **68**, 521.
VAN DER LEEDEN, P., 1955, *Physica*, **21**, 545.
LYPE, E. F., 1946, *Phys. Rev.*, **69**, 652.
PIPPARD, A. B., 1956, *Physica*, **22**, 99.
PRICE, P. J., 1951, *J. Chem. Phys.*, **19**, 1281.
SHOENBERG, D., 1952, *Superconductivity* (Cambridge : University Press).
TEMPERLEY, H. N. V., 1956, *Changes of State* (London : Cleaver-Hume Press).
ZIMM, B. H., 1951, *J. Chem. Phys.*, **19**, 1019.

A Statistical Distribution arising in the Study of the Ionosphere

By M. S. LONGUET-HIGGINS
National Institute of Oceanography, Wormley, Surrey

MS. received 7th February 1957

Abstract. An approximate distribution is deduced for the directions of the 'lines of maxima' on a random surface. The theoretical distribution is found to be in good agreement with some experimental results obtained previously by Briggs and Page.

---

§ 1. INTRODUCTION

THE analysis of fading patterns of radio waves reflected from the ionosphere gives rise to some interesting statistical problems. Among these is one studied by Briggs and Page (1955) which may be stated as follows.

Suppose we are given a two-dimensional surface whose height represents, say, the intensity $f(x, y)$ of a beam of reflected radiation as a function of two horizontal coordinates x and y. Let two parallel sections of the surface be taken, along the lines $y = y_0$ and $y = y_0 + \delta$. Consider the maxima of the sections along these lines. If the lines are sufficiently close (δ sufficiently small) the maxima of the two functions will correspond in pairs, the members of a pair being slightly shifted relative to one another. If the line joining a pair of maxima makes an angle α with the y-axis the problem is to find the statistical distribution of $\tan \alpha$.

Lacking a theoretical solution, Briggs and Page constructed pairs of random sequences to represent the functions $f(x, y_0)$ and $f(x, y_0 + \delta)$. These sequences were arranged so as to be normally distributed with zero mean and unit standard deviation. Further, their auto-correlograms and cross-correlograms were also of normal form. The observed distribution of $\tan \alpha$ for the pairs of sequences is indicated by the experimental points in the figure, which are transcribed from Fig. 3 of Briggs and Page (1955).

The theoretical problem, however, is similar to many that have been treated in connection with the analysis of the sea surface (see Longuet-Higgins 1957 a, b) and in this note we shall present a solution. It will be shown that, provided only the joint distribution of $f(x, y)$ and its derivatives is normal (a condition satisfied in very general circumstances) the probability distribution of $\tan \alpha$ is given by

$$p(\tan \alpha) = \frac{A^2}{2[A^2 + (\tan \alpha - B)^2]^{3/2}}, \qquad \ldots\ldots(1)$$

where A and B are constants which depend only on the form of the correlogram of f. When the surface is isotropic, so that its correlogram has circular symmetry, we have

$$A^2 = 1/3, \qquad B = 0$$

and so

$$p(\tan \alpha) = \frac{1}{6[\frac{1}{3} + \tan^2 \alpha]^{3/2}}. \qquad \ldots\ldots(2)$$

In the case treated by Briggs and Page we have apparently (see § 5 below)

$$A^2 = 2/3, \qquad B = 0$$

and so

$$p(\tan\alpha) = \frac{1}{3[\frac{2}{3} + \tan^2\alpha]^{3/2}}. \qquad \ldots\ldots(3)$$

This curve is shown by the dotted line in the figure. It will be seen that the agreement is fairly close. Moreover, the curve is in adequate agreement at the tails of the distribution, in contrast with the normal frequency curve tentatively suggested by Briggs and Page.

§ 2. THE DERIVATION OF $p(\tan\alpha)$

Suppose that two corresponding maxima are separated by intervals dx and dy in the x and y directions. Then, since the gradient $\partial f/\partial x$ vanishes at both maxima we have

$$\frac{\partial f}{\partial x}(x+dx, y+dy) - \frac{\partial f}{\partial x}(x, y) = 0$$

or

$$\frac{\partial^2 f}{\partial x^2}dx + \frac{\partial^2 f}{\partial x \partial y}dy = 0$$

approximately, and therefore

$$\tan\alpha = \frac{dx}{dy} = -\frac{\partial^2 f/\partial x\partial y}{\partial^2 f/\partial x^2}. \qquad \ldots\ldots(4)$$

Clearly this formula also holds for any pair of points having the *same* gradient not necessarily zero.

Now, let us assume that f is a stationary random function whose partial derivatives up to the second order are distributed normally.

Writing for convenience

$$\frac{\partial f}{\partial x}, \frac{\partial^2 f}{\partial x^2}, \frac{\partial^2 f}{\partial x \partial y} = \xi_1, \xi_2, \xi_3$$

we have for the probability-density of (ξ_1, ξ_2, ξ_3)

$$p(\xi_1, \xi_2, \xi_3) = \frac{1}{(2\pi)^{3/2}|\Xi_{ij}|^{1/2}} \exp[-\tfrac{1}{2}\sum_{i,j} M_{ij}\xi_i\xi_j], \qquad \ldots\ldots(5)$$

where (M_{ij}) is the reciprocal of the matrix of mean values

$$(\Xi_{ij}) = (\overline{\xi_i\xi_j}).$$

In fact, if $\psi(X, Y)$ denotes the correlation function

$$\psi(X, Y) \equiv \overline{f(x, y)f(x+X, y+Y)} \qquad \ldots\ldots(6)$$

we have by repeated differentiation

$$\left.\begin{aligned} \overline{\xi_1^2} &= -\frac{\partial^2\psi}{\partial X^2}(0,0) = m_{20}, \quad \text{say}, \\ \overline{\xi_2^2} &= \frac{\partial^4\psi}{\partial X^4}(0,0) = m_{40}, \\ \overline{\xi_3^2} &= \frac{\partial^4\psi}{\partial X^2 \partial Y^2}(0,0) = m_{22}, \end{aligned}\right\} \qquad \ldots\ldots(7)$$

and also

$$\left.\begin{aligned} \overline{\xi_2\xi_3} &= \frac{\partial^4\psi}{\partial X^3\partial Y}(0,0) = m_{31} \\ \overline{\xi_1\xi_2} &= \overline{\xi_1\xi_3} = 0. \end{aligned}\right\} \qquad \ldots\ldots(8)$$

So

$$(\Xi_{ij}) = \begin{bmatrix} m_{20} & 0 & 0 \\ 0 & m_{40} & m_{31} \\ 0 & m_{31} & m_{22} \end{bmatrix}$$

and

$$p(\xi_1, \xi_2, \xi_3) = p(\xi_1)\, p(\xi_2, \xi_3), \qquad \ldots\ldots(9)$$

where

$$\left.\begin{aligned} p(\xi_1) &= \frac{1}{(2\pi m_{20})^{1/2}} \exp(-\xi_1{}^2/2m_{20}), \\ p(\xi_2, \xi_3) &= \frac{1}{(2\pi\Delta)^{1/2}} \exp[-(m_{22}\xi_2{}^2 - 2m_{31}\xi_2\xi_3 + m_{40}\xi_3{}^2)/2\Delta], \\ \Delta &= m_{40}m_{22} - m_{31}{}^2 \end{aligned}\right\} \qquad \ldots\ldots(10)$$

$p(\xi_1)$ and $p(\xi_2, \xi_3)$ are the distributions of ξ_1 and of (ξ_2, ξ_3) separately, and equation (9) shows that they are statistically independent.

We wish now to find the distribution of the quantity

$$\eta = \tan\alpha = -\xi_3/\xi_2$$

at the points where the gradient ξ_1 vanishes. First let us find the joint distribution of (ξ_2, ξ_3) at such points; we shall denote this by $p(\xi_2, \xi_3)_{\xi_1}$. Now since ξ_2 and ξ_3 are statistically independent of ξ_1, the distribution of (ξ_2, ξ_3) for points on the x-axis such that $0 < \xi_1 < d\xi_1$ is the same as the ordinary distribution $p(\xi_2, \xi_3)$ for points distributed randomly and uniformly with regard to x. But near any particular zero of ξ_1 the function ξ_1 remains in the interval $0 < \xi_1 < d\xi_1$ for a distance dx inversely proportional to $|\partial\xi_1/\partial x| = |\xi_2|$. Hence we have

$$p(\xi_2, \xi_3) = C|\xi_2|^{-1}\, p(\xi_2, \xi_3)_{\xi_1}$$

or

$$p(\xi_2, \xi_3)_{\xi_1} = C^{-1}|\xi_2|\, p(\xi_2, \xi_3),$$

where C is a normalizing constant. This gives

$$p(\xi_2, \xi_3)_{\xi_1} = \frac{1}{2(2\pi\Delta m_{40})^{1/2}} |\xi_2| \exp[-m_{22}\xi_2{}^2 - 2m_{31}\xi_2\xi_3 + m_{40}\xi_3{}^2/2\Delta]. \qquad \ldots\ldots(11)$$

To find the distribution of η we now put

$$(-\xi_3/\xi_2, \xi_2) = (\eta, \eta'), \qquad \frac{\partial(\xi_2, \xi_3)}{\partial(\eta, \eta')} = \eta',$$

giving

$$p(\eta, \eta')_{\xi_1} = \frac{1}{2(2\pi\Delta m_{40})^{1/2}} \eta'^2 \exp[-\eta'^2(m_{22} + 2m_{31}\eta + m_{40}\eta^2)/2\Delta]. \qquad \ldots\ldots(12)$$

On eliminating η' by integration over the whole range $(-\infty, \infty)$ we find

$$p(\eta)_{\xi_1} = \frac{\Delta/m_{40}{}^{1/2}}{2[m_{22} + 2m_{31}\eta + m_{40}\eta^2]^{3/2}}$$

or

$$p(\eta)_{\xi_1} = \frac{1}{2[A^2 + (\eta - B)^2]^{3/2}}, \qquad \ldots\ldots(13)$$

where

$$A = \frac{\Delta^{1/2}}{m_{40}}, \qquad B = -\frac{m_{31}}{m_{40}}. \qquad \ldots\ldots(14)$$

This is the result stated in equation (1).

We have included in this distribution all points at which the gradient vanishes, i.e. both maxima and minima. However, since $p(\eta, \eta')_{\xi_1}$ is symmetrical about the origin, the contributions to $p(\eta)_{\xi_1}$ from the maxima ($\eta' < 0$) are equal to the contributions from the minima ($\eta' > 0$). So, after introducing the appropriate normalizing constant, we see that equation (1) represents the distribution of $\tan\alpha$ for the maxima alone, and also for the minima.

Clearly $p(\eta)_{\xi_1}$ is a symmetrical distribution centred on $\eta = B$ and falling off like η^{-3} for large values of η. Thus the standard deviation is infinite. But a convenient measure of the width of the distribution is the interquartile range, which is

$$\frac{2}{\sqrt{3}} A = 1{\cdot}155 \ldots A. \qquad \ldots\ldots(15)$$

The width of the distribution at a level equal to half its maximum height is

$$2\sqrt{(4^{1/3} - 1)} A = 1{\cdot}532 \ldots A. \qquad \ldots\ldots(16)$$

§ 3. REMARK

We have seen that the angle α can be defined at all points of the curve, being the angle made with the y-axis by lines of constant $\partial f / \partial x$. The distribution of $\tan\alpha$ for points where $\partial f / \partial x$ takes any particular value ξ_1 other than zero is also given by equation (1); this follows from the fact that $p(\xi_2, \xi_3)_{\xi_1}$ is independent of ξ_1 itself.

However, if instead of the distribution of $\tan\alpha$ at points of given gradient we wish to find the distribution of $\tan\alpha$ at points randomly and uniformly distributed with regard to x, then we have only to replace $p(\xi_2, \xi_3)_{\xi_1}$ in the foregoing analysis by the ordinary distribution $p(\xi_2, \xi_3)$. As a consequence we find the following distribution for $\tan\alpha$:

$$p^{(1)}(\tan\alpha) = \frac{1}{\pi} \frac{1}{A^2 + (\tan\alpha - B)^2}, \qquad \ldots\ldots(17)$$

where A and B have their former meanings. This expression is similar in form to the previous distribution, having a mean value B and a width proportional to A. However, the two distributions are not to be confused, and in fact the former distribution (1) is the one which is appropriate in the present case, considering the method of observing the data. For a further discussion of these distributions, which arise also in the analysis of moving waveforms, the reader is referred to Longuet-Higgins (1957 b).

§ 4. AN ISOTROPIC SURFACE

Assuming that the derivatives of the correlation function $\psi(X, Y)$ at the origin exist up to the fourth order, we have in general

$$\begin{aligned} \psi(X, Y) = C_{00} &+ (C_{10}X + C_{01}Y) \\ &+ (C_{20}X^2 + C_{11}XY + C_{02}Y^2) \\ &+ (C_{30}X^3 + C_{21}X^2Y + C_{12}XY^2 + C_{03}Y^3) \\ &+ (C_{40}X^4 + C_{31}X^3Y + C_{22}X^2Y^2 + C_{13}XY^3 + C_{04}Y^4) \\ &+ \mathrm{R} \end{aligned}$$

where

$$C_{pq} = \frac{1}{p!\,q!} \frac{\partial^{p+q}\psi}{\partial X^p \partial Y^q}(0, 0)$$

and R is a remainder term of higher order. Suppose now that the surface is isotropic. Then from the two conditions

$$\left.\begin{aligned}\psi(X,Y)&=\psi(-X,Y)\\ \psi(X,Y)&=\psi(X,-Y)\end{aligned}\right\}$$

we see that C_{pq} vanishes whenever p is odd and whenever q is odd, respectively. The remaining terms in the expansion of ψ are integral powers of both X^2 and Y^2. Further, since the surface is isotropic, $\psi(X, Y)$ is a function of X^2+Y^2 only. The only polynomial of the $2n$th degree having this form is $(X^2+Y^2)^n$. Hence

$$\psi(X,Y)=D_0+D_1(X^2+Y^2)+D_2(X^4+2X^2Y^2+Y^4)+R,$$

where D_0, D_1 and D_2 are constants. This gives

$$\left.\begin{aligned}m_{40}&=\frac{\partial^4\psi}{\partial X^4}(0,0)=24D_2\\ m_{31}&=\frac{\partial^4\psi}{\partial X^3\partial Y}(0,0)=0\\ m_{22}&=\frac{\partial^4\psi}{\partial X^2\partial Y^2}(0,0)=8D_2\end{aligned}\right\}$$

and so

$$A^2=\frac{\Delta}{m_{40}^2}=\frac{1}{3};\qquad B=0. \qquad\ldots\ldots(18)$$

With these values of A and B the distribution (13) becomes

$$p(\tan\alpha)=\frac{1}{6[\frac{1}{3}+\tan^2\alpha]^{3/2}}. \qquad\ldots\ldots(19)$$

Thus $p(\tan\alpha)$ has a mean value zero and an interquartile range 2/3. The width of the distribution at half the maximum level is 0·884....

The conditions (18), although necessary for an isotropic surface, are not sufficient. For example, a surface consisting of three long-crested but incoherent wave systems whose directions make angles 0, $2\pi/3$, $4\pi/3$ with a fixed direction would also satisfy the same conditions. The surface would therefore give rise to the same distribution of $\tan\alpha$ as does an isotropic surface.

§ 5. THE SURFACE STUDIED BY BRIGGS AND PAGE (1955)

To represent the two functions $f(x, y_0)$ and $f(x, y_0+\delta)$ Briggs and Page first constructed two sequences of numbers

$$P_i=\sum_{j=-\infty}^{\infty}a_j\xi_{i+j};\qquad Q_i=\sum_{j=-\infty}^{\infty}a_j\eta_{i+j}, \qquad\ldots\ldots(20)$$

where ξ_i and η_i ($i=\ldots-2,-1,0,1,2,\ldots$) were two sets of mutually independent random normal deviates (each with zero mean and unit standard deviation) and where the a_i were fixed weighting constants given by

$$a_j=\text{constant}\times\exp(-j^2/\sigma^2). \qquad\ldots\ldots(21)$$

It can be shown that under these circumstances each of the sequences P_iQ_i is normally distributed, and further that their auto-correlations and cross-correlations are given by

$$\left.\begin{aligned}&\overline{P_iP_{i+n}}=\overline{Q_iQ_{i+n}}=\text{constant}\times\exp(-n^2/2\sigma^2),\quad (n\text{ even})\\ &\overline{P_iQ_{i+n}}=0.\end{aligned}\right\}\ldots\ldots(22)$$

The constant in (22) is taken to be unity. The authors then represent $f(x, y_0)$ by the series P_i and $f(x, y_0+\delta)$ by the series

$$R_i = \rho P_i + (1-\rho^2)^{1/2} Q_i,$$

where ρ is a constant, equal to $\exp(-\delta^2/\sigma^2)$. It can easily be seen that

$$\left.\begin{aligned} \overline{R_i R_{i+n}} &= \exp(-n^2/2\sigma^2), \\ \overline{P_i R_{i+n}} &= \exp(-n^2/2\sigma^2). \end{aligned}\right\}$$

Thus

$$\left.\begin{aligned} \overline{f(x, y_0) f(x+n, y_0)} &= \exp(-n^2/2\sigma^2), \\ \overline{f(x, y_0+\delta) f(x+n, y_0+\delta)} &= \exp(-n^2/2\sigma^2), \\ \overline{f(x, y_0) f(x+n, y_0+\delta)} &= \exp[-(n^2+2\delta^2)/2\sigma^2], \end{aligned}\right\} \qquad \ldots\ldots(23)$$

so that, in so far as the two functions are concerned, they behave like sections of a surface having a correlation function

$$\psi(X, Y) = \exp[-(X^2+2Y^2)/2\sigma^2]. \qquad \ldots\ldots(24)$$

For this surface we have clearly

$$\left.\begin{aligned} m_{40} &= \frac{\partial^4 \psi}{\partial X^4}(0,0) = \frac{3}{\sigma^4}, \\ m_{31} &= \frac{\partial^4 \psi}{\partial X^3 \partial Y}(0,0) = 0, \\ m_{22} &= \frac{\partial^4 \psi}{\partial X^2 \partial Y^2}(0,0) = \frac{2}{\sigma^4}, \end{aligned}\right\} \qquad \ldots\ldots(25)$$

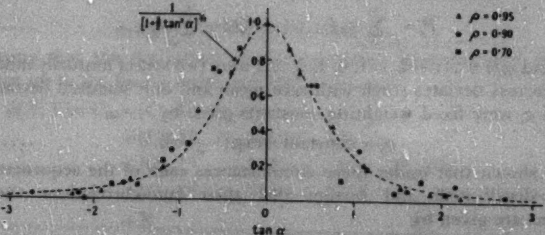
giving

$$A^2 = 2/3, \qquad B = 0.$$

The corresponding curve is

$$p(\tan\alpha) = \frac{1}{3[\frac{2}{3} + \tan^2\alpha]^{3/2}} = \frac{(3/8)^{1/2}}{[1 + \frac{3}{2}\tan^2\alpha]^{3/2}}. \qquad \ldots\ldots(27)$$

The experimental points of Briggs and Page are shown in the figure. These were not normalized, but were reduced so that the maximum ordinate, for each value of ρ, was equal to unity. The broken curve in the figure corresponds to the distribution (27) multiplied by a factor $(8/3)^{1/2}$ to bring the maximum ordinate also to unity. It will be seen that there is reasonably good agreement.



The theoretical curve of equation (27) (with vertical scale multiplied by $(8/3)^{1/2}$) compared with the experimental points of Briggs and Page.

It should be noted that in their paper Briggs and Page assumed that $P_i P_{i+n}$ was proportional to $\exp(-n^2/\sigma^2)$, omitting the factor 2 in equation (22)†. Consequently, they assumed the correlation function to be of the form

$$\psi(X, Y) = \exp[-(X^2+Y^2)/2\sigma^2],$$

which would represent an isotropic surface. In fact an isotropic correlogram could have been obtained by assuming

$$\rho = \exp(-\delta^2/\sigma^2)$$

and by plotting the frequency curves in Fig. 3 of their paper against the parameter $(\Delta/\sigma)(-2\log\rho)^{1/2}$ instead of against $(\Delta/\sigma)(-\log\rho)^{1/2}$.

§ 6. CONCLUSIONS

The distribution of $\tan\alpha$ for a random, normal surface is in general given by equation (13), where $\eta = \tan\alpha$ and A and B are constants depending on the correlation function of the surface. This is in agreement with the experimental distribution found by Briggs and Page.

For an isotropic surface the distribution of $\tan\alpha$ is given by equation (19).

ACKNOWLEDGMENT

I am indebted to Mr J. A. Ratcliffe for drawing my attention to this problem.

REFERENCES

BRIGGS, B. H., and PAGE, E. S., 1955, *Report of the* 1954 *Conference on The Physics of the Ionosphere* (London : Physical Society), p. 119.

LONGUET-HIGGINS, M. S., 1957 a, *Phil. Trans. Roy. Soc.* A, **249**, 321; 1957 b, *Proc. Camb. Phil. Soc.*, **53**, 23.

---

† Dr. Briggs has kindly confirmed this in a personal communication.

The Magnetic Susceptibilities of Lanthanum, Cerium, Praseodymium, Neodymium and Samarium, from 1·5°K to 300°K

By J. M. LOCK
Royal Radar Establishment, Great Malvern, Worcs.

MS. received 10*th January* 1957, *and in revised form* 25*th February* 1957

Abstract. The magnetic susceptibilities of the rare-earth metals lanthanum, cerium, praseodymium, neodymium and samarium have been measured over the temperature range 1·5°K to 300°K. Low temperature anomalies correlating with recent specific heat measurements have been found in cerium, neodymium and samarium at temperatures of 12·5°K, 7·5°K and 14·8°K respectively, and it is believed that these metals become anti-ferromagnetic below these temperatures. At higher temperatures the susceptibilities of cerium and praseodymium obey the Curie–Weiss law, with effective atomic moments of 2·51 and 3·56 Bohr magnetons respectively. Neodymium obeys a law of the form $\chi = C/(T-\theta)+K$, where the Curie constant C indicates an effective atomic moment of 3·3 Bohr magnetons (μ_B). These moments are in fair agreement with the values expected for the trivalent ions of the elements, assuming Russell–Saunders coupling of the orbital and spin moments of the 4f electrons.

§ 1. Introduction

The magnetic properties of the salts of the rare earths have been extensively studied, so that the behaviour of the 4f shell of electrons, which is responsible for the strong paramagnetism of these substances, under conditions of high magnetic dilution, is well known. On the other hand magnetic measurements on the pure metals themselves had previously been restricted mainly to temperatures above about 80°K, so that very little was known about the magnetic interactions which might be expected to occur at lower temperatures. The present investigation was undertaken in order to learn more about such interactions, particularly in the elements in the first half of the rare-earth group.

A preliminary account of some of this work was given at the Conférence de Physique des Basses Températures, Paris 1955.

§ 2. Experimental Details

2.1. *Apparatus*

In view of the comparatively large susceptibilities of the rare-earth metals, the sensitivity required for their measurement is not great, and a Sucksmith balance was used to measure the force on a small specimen in an inhomogeneous magnetic field provided by a suitable electromagnet.

A null method was used in which the force on the specimen could be balanced by the force between one moving and two fixed coils connected up as an Ampère balance. The deflection of the beam of light from the balance was amplified by a split photocell connected to a long-period galvanometer, and the sensitivity was about 4 cm on the galvanometer scale for a force of 1 mg on the specimen. Under optimum conditions the 'noise level' was about 1 mm, and in addition there was usually a zero drift of a few millimetres per minute.

The specimen was suspended in an atmosphere of helium exchange gas in a copper tube 1 cm in internal diameter surrounded by a vacuum chamber. This was immersed in the cooling bath of liquid helium, hydrogen or nitrogen contained in a glass Dewar, which fitted between the pole pieces of the electromagnet. Two heating coils wound on the copper tube enabled it to be heated over a suitable range of temperatures above that of the cooling bath, and the temperature of the specimen was measured by resistance thermometers (platinum for temperatures down to 20°K, and carbon for lower temperatures) in the exchange gas just below it. These thermometers were calibrated against the vapour pressure of helium, hydrogen, oxygen and nitrogen and a suitable interpolation made between these ranges. Except in the ranges covered by the vapour pressure of helium and hydrogen, the errors in temperature measurement are estimated to be about ±0·05°K up to 20°K and ±0·2°K above 20°K.

For most of the measurements a small electromagnet giving fields up to about 5000 oersteds was used. The pole-piece faces were machined to a cylindrical shape, with the axis of the cylinder horizontal, to give a roughly calculable field distribution. With this design the lines of force and the magnetic equipotentials are approximately coaxial circles, and the variation of magnetic force on an infinitesimal specimen along the symmetry axis can readily be calculated. By choosing the radius of curvature and separation of the pole pieces appropriately, the force near its maximum can then be made uniform over a suitable volume to any required degree of accuracy. In practice the force dropped off by 2% at points 0·5 cm above and below its maximal position. Since the specimens used were about 6 mm in length the correction to be applied for this finite size is about 0·2%.

The current supply for the magnet is provided by a motor generator, and an electronic servo system enables the current to be controlled and stabilized to at least 1 part in 10^4. The magnetic field was calibrated in the usual way with a search-coil, ballistic galvanometer, and standard mutual inductance, but absolute values of susceptibility were obtained more accurately by comparison with freshly prepared single crystals of chrome alum, using the results of de Haas and Gorter (1930). It is estimated that relative values of susceptibility are accurate to about 0·5%, and absolute values to about 2%.

2.2. *Specimens*

The metals which have been studied are cerium, praseodymium, neodymium, samarium and lanthanum. Apart from one sample of cerium from the Atomic Energy Research Establishment, Harwell, all the metals have been supplied by Johnson, Matthey and Co. As the presence of iron is particularly undesirable a careful analysis of the quantity present was obtained, and is quoted together with other details of the analysis where it is known. The specimens used for the susceptibility measurements were roughly rectangular lumps, cut from the polycrystalline ingots with a clean hacksaw, and polished with fine sandpaper; they were then mounted with thin paper and Durofix on a thin quartz rod which was afterwards fused to the specimen suspension. This was then placed in the apparatus, and the specimen chamber was evacuated and filled with clean helium exchange gas. In this way very little oxidation took place and no observable change in the specimen properties was noticed over a period of sometimes a few weeks when left in the apparatus.

2.3. *Method of Analysis of Results*

The force on a body of magnetic moment M in an inhomogeneous magnetic field H is given by the equation

$$F = M(\partial H/\partial z) \qquad \ldots\ldots(1)$$

where H is varying in the vertical or z direction.

In the case of a pure isotropic paramagnetic material,

$$M = m\chi H \qquad \ldots\ldots(2)$$

where m is the mass of the specimen and χ its mass susceptibility, so that we have the usual relation

$$F = m\chi H(\partial H/\partial z). \qquad \ldots\ldots(3)$$

By making measurements on a freshly prepared crystal of chrome alum, and using the field values obtained from the search-coil calibration, accurate values of $\partial H/\partial z$ can be derived using this equation.

If, however, the material is not a pure paramagnetic, or if there is any ferromagnetic impurity in it we can use (1) to obtain values of M in different applied fields. In the latter case it is found in practice that a plot of M against H gives a straight line above fairly low fields, suggesting that the ferromagnetic impurity then becomes saturated. The slope of this line is then equal to $m\chi$ for the specimen. The intercept M_0 of the straight line on the axis of M can also be obtained and in the case of a ferromagnetic impurity gives a measure of its saturation moment. In most of this work it has been considered sufficient to calculate χ and M_0 from measurements in two different fields (about 4800 and 2500 oersteds). At very low temperatures some of the rare-earth specimens exhibit intrinsic non-linear magnetic effects; usually the plot of M against H is a straight line above quite small fields and in this case the value of M_0 gives a good indication of the magnitude of the effect.

In the case of weakly magnetic specimens a correction has been applied for the slight magnetic effect of the suspension, determined separately.

§ 3. EXPERIMENTAL RESULTS

3.1. *Lanthanum*

The lanthanum specimen measured approximately 10 mm × 5 mm × 5 mm and weighed 1·36 g. The main impurities were said to be Li 0·05% and Fe 0·004%. The crystal structure at room temperature was investigated using an X-ray tube with a chromium target; separate test specimens of this and the other metals were prepared in the form of thin rods about ½ mm in diameter for this purpose. Powder photographs were taken, and the lanthanum proved to be predominantly hexagonal close-packed, but with some fainter lines corresponding to face-centred cubic structure.

The susceptibility was measured from room temperature down to about 6°K. Faint traces of superconductivity appeared at 5·4°K which prevented measurements of the normal susceptibility from being carried out below this temperature. At 4·2°K the diamagnetic volume susceptibility was $1{\cdot}2/4\pi$ in a field of 300 oersteds, suggesting that the bulk of the specimen was in a superconducting state. The paramagnetic susceptibility above 6°K is plotted against temperature in figure 1. Trombe (1934) found a value of χ at room temperature of about $1{\cdot}2\times10^{-4}$, some 60% higher than our value. Possibly the iron content of his specimen was greater than ours.

3.2. *Cerium*

Specimens of cerium from two sources were studied, one from Harwell containing 0·001% iron and the other from Johnson, Matthey and Co. containing 0·03% iron. Both originally showed x-ray lines at room temperature corresponding to face-centred cubic structure only, but one, which was used for specific heat measurements as well as magnetic measurements and which was eventually cooled down to 20°K from room temperature 50 times, showed additional faint lines corresponding to hexagonal close-packed structure after this treatment. The significance of this will be discussed later.

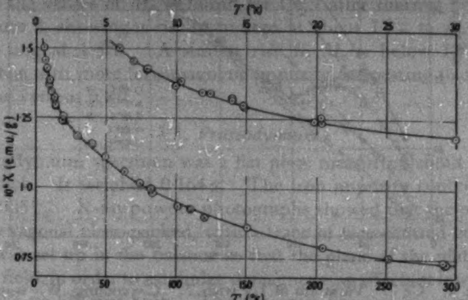


Figure 1. La 1. Variation of χ with T.

One specimen from each of these two sources was studied in detail, after some preliminary measurements on earlier specimens had shown that the low-temperature behaviour of cerium depends strongly on its previous thermal history. The specimens were roughly 4 mm cubes weighing about 0·3–0·4 g, and their susceptibilities were measured from room temperature down to 1·5°K. The results for the two specimens of different purity were very similar, except that the material with the larger amount of iron impurity exhibited a much larger value of M_0/m (where m is the mass of the specimen), a maximum value of about 5×10^{-2} e.m.u./g being attained at about 100°K, while the purer specimen at the same temperature had an M_0/m of only 3×10^{-3} e.m.u./g. The reciprocal of the susceptibility per gramme of the purer specimen, Ce 7, is plotted against temperature in figures 2 and 3. The susceptibility values for the less pure specimens are about 7% higher over the whole temperature range.

Between about 40°K and 200°K the susceptibility of cerium shows pronounced thermal hysteresis, as previously reported by Trombe (1934) and La Blanchetais (1945) in measurements extending down to 80°K, the susceptibility on warming up being less than on cooling. This hysteresis loop gradually disappears however on repeated runs through this temperature range, and figure 2 shows that after about one hundred cycles of temperature the area of the loop is reduced to only about 5% of its initial value. To achieve this large number of coolings the specimen was removed from the apparatus and sealed in a small glass tube with a little helium exchange gas, and was then dipped in a liquid hydrogen bath and warmed up to room temperature the appropriate number of times.

Along the reversible part of the cooling curves from room temperature down to about 110°K the susceptibility of the purer specimen obeys a Curie–Weiss law

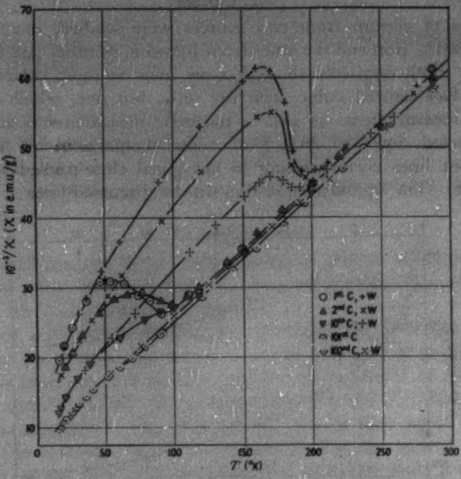


Figure 2. Ce 7. Variation of $1/\chi$ with T. C=cooling, W=warming.

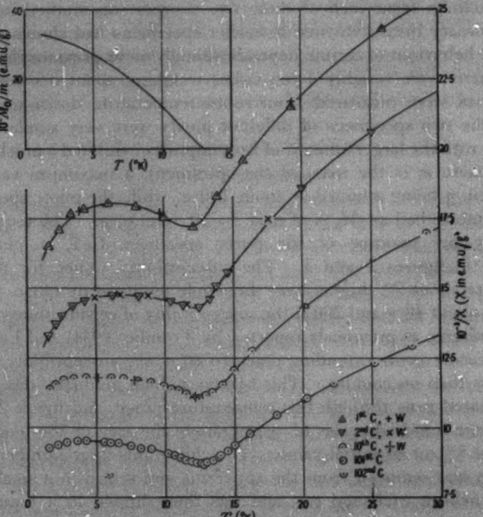


Figure 3. Ce 7. Variation of $1/\chi$ with T. C=cooling, W=warming. Inset: smoothed values of M_0/m.

$\chi = C/(T-\theta)$ where the Curie constant C has the same value of $55{\cdot}3 \times 10^{-4}$ e.m.u./g for all coolings, but the Weiss constant θ varies from -46°K for the initial cooling to -38°K for the 100th cooling. The value of the Curie constant corresponds to an effective atomic moment of $2{\cdot}51\mu_B$, close to that predicted for the free ion Ce^{3+} ($2{\cdot}54\mu_B$). These values are in reasonable agreement with those found by La Blanchetais (1945).

Figure 3 shows that in the low-temperature range the susceptibility of cerium reaches a maximum between 12°K and 13°K, thereafter falling off slightly, but rising again below about 4°K. Below 12°K non-linear magnetic effects also begin to occur, and the values of M_0/m found for Ce 7 after thermal cycling are also plotted in figure 3, showing a rise from zero at about 13°K to a value of about 33×10^{-3} e.m.u./g at 1·5°K. A similar rise in M_0/m below 13°K was found in the specimens with more ferromagnetic impurity, suggesting that it is an effect inherent in the cerium itself.

3.3. *Praseodymium*

The praseodymium specimen was a flat plate measuring about 4 mm × 4 mm and 1½ mm thick. It weighed 0·164 g. The iron impurity content was said to be less than 0·005%. X-ray powder photographs showed that the praseodymium was mainly hexagonal close-packed, with a trace of face-centred cubic structure present. It was set up in the balance so that the plane of the plate was parallel to the applied field, in order to minimize demagnetizing effects at low temperatures when the susceptibility was high. The susceptibility was measured from room temperature down to 1·3°K. Figure 4 shows the values of $1/\chi$ plotted against temperature. The maximum 'ferromagnetic correction' to the susceptibility was about 2% at 11°K where M_0/m rose to a maximum value of about 0·13 e.m.u./g.

Above about 100°K χ obeys a Curie law, $\chi = (11{\cdot}25/T) \times 10^{-3}$ e.m.u./g, from which the effective atomic moment is calculated to be $3{\cdot}56\mu_B$, compared with a predicted value of $3{\cdot}58\,\mu_B$ for the free ion Pr^{3+}. Below 100°K the susceptibility

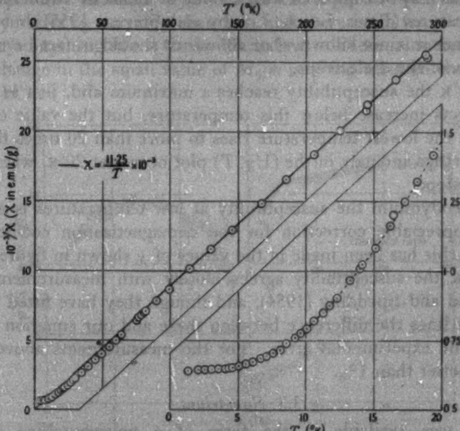


Figure 4. Pr 1. Variation of $1/\chi$ with T.

first increases above the Curie law value and then flattens out at a value of $1{\cdot}55\times10^{-3}$ e.m.u./g below about 4°K. For such a high susceptibility there is an appreciable correction for the demagnetization factor which can be calculated roughly from the formula $1/K=1/K_0-4\pi n$ where K_0 is the observed volume susceptibility ($=\rho\chi$), K the true volume susceptibility, and $4\pi n$ the demagnetization factor for the specimen. An approximate formula for a thin disc of diameter d and thickness t, with the field in the plane of the disc, is $4\pi n=\pi^2t/d$ so that, for our specimen, very approximately $4\pi n\sim4$. Thus we have $1/\chi=1/\chi_0-4\pi n\rho$, and the maximum correction in the case of praseodymium is about 4%. This correction has been applied to the values of $1/\chi$ shown in figure 4.

3.4. *Neodymium*

Two specimens of neodymium were measured, weighing 0·306 and 0·080 g, and measuring about 4 mm × 5 mm × 2·8 mm and 3·5 mm × 2·8 mm × 1·4 mm, respectively. The main impurities were said to be Fe 0·03% and Li 0·4%. X-ray powder photographs indicated that the material was predominantly hexagonal close-packed, with some face-centred cubic structure present. The experimental values of $1/\chi$ and smoothed values of M_0/m for the first specimen are plotted against temperature in figure 5. Above 35°K the susceptibility of this specimen is given to within 1% by the equation

$$\chi=\frac{9{\cdot}47\times10^{-3}}{T-4{\cdot}3}+5{\cdot}0\times10^{-4}\text{e.m.u./g.}$$

The Curie constant $9{\cdot}47\times10^{-3}$ corresponds to an effective atomic moment of $3{\cdot}3\,\mu_B$, compared with the theoretical value of $3{\cdot}62\,\mu_B$ for the free Nd^{3+} ion. The constant term in the susceptibility may be ascribed partly to the paramagnetism of the d electrons (cf. lanthanum) and partly to the proximity of the ground state ($j=9/2$) to the first excited state ($j=11/2$) of the 4f electrons. An energy gap of the order of 630 cm^{-1} is small enough to account approximately for the effect. This may be compared with a value of 1350 cm^{-1} from measurements at high temperatures (Bates, Leach, Loasby and Stevens 1955), but both values are suspect, since it is not known what allowance should in fact be made for the paramagnetism of the d electrons.

At about 8°K the susceptibility reaches a maximum and, just as for cerium, non-linear effects increase below this temperature, but the value of M_0/m for neodymium at the lowest temperature rises to more than 20 times the value for cerium. A further anomaly in the $(1/\chi, T)$ plot occurs at 20°K, where there is a change in the slope.

As in praseodymium the susceptibility at low temperatures is so high that there is an appreciable correction for the demagnetization coefficient of the specimen, and this has been made in the values of χ shown in figure 5.

Above 20°K the susceptibility agrees closely with measurements made by Elliott, Legvold and Spedding (1954), and though they have fitted their points to two straight lines the difference between these and our equation is probably smaller than the experimental error. For the measurements above 145°K the agreement is better than 1%.

3.5. *Samarium*

Samarium was available in the form of a polycrystalline agglomerate condensed from the vapour on a thin tantalum sheet. A specimen was cut from

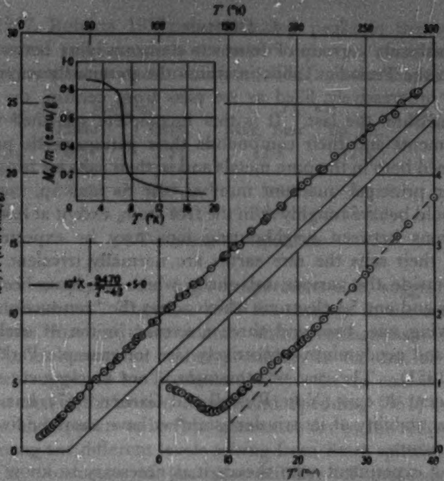


Figure 5. Nd 1. Variation of $1/\chi$ with T. Inset: smoothed values of M_0/m.

this measuring roughly 11 mm × 4 mm and of irregular thickness. Its mass was 0·273 g. The main impurities were said to be Fe 0·006% and Ca 0·002%.

The measured values of $1/\chi$ are plotted against temperature in figure 6. The main features are the maximum in the susceptibility at 15°K, and the sharp drop just below that temperature. Below 7°K the susceptibility starts to rise again, and between 110°K and 150°K there is a kink in the curve corresponding to a drop in the susceptibility of about 5%. From 150°K up to room temperature the susceptibility varies by no more than 2%, and there is no temperature range in which it obeys anything like a Curie–Weiss law. There is no indication of any significant change in the small value of M_0/m in going through the susceptibility anomaly at 15°K.

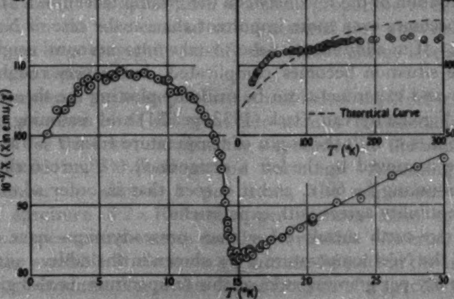


Figure 6. Sm 1. Variation of $1/\chi$ with T.

§ 4. Discussion

The rare earths are a group of fourteen elements lying between lanthanum and hafnium in the Periodic Table, in which the previously empty 4f sub-shell of electrons is progressively filled as we pass from cerium, the first of these elements, to lutetium, the last. It is this incomplete sub-shell which confers upon these elements and their compounds their paramagnetic properties, and since it is screened both in the pure metals and in their salts by two complete sub-shells of higher principal quantum number, the 5s and 5p, containing eight electrons, it should behave roughly as in the free atom, except at low temperatures where interactions between neighbouring ions may be expected to become important. In their salts the rare earths are normally trivalent, and we shall suppose that outside the various sub-shells which we have been considering there are two 6s and one 5d electrons which act as the 'conduction' electrons in the metals, leaving one, two, and three electrons in the 4f shells of cerium, praseodymium and neodymium respectively (see for example Parkinson, Simon and Spedding 1951). The case of samarium is not so clear cut since divalent salts of this element do exist (Yost, Russell and Garner 1947), but on account of the much greater stability of its trivalent state we have assumed five 4f electrons for metallic samarium.

In comparing experiment with theory it is necessary to know the values of the quantum numbers L, S, J in order to calculate the effective atomic moment, μ_{eff} Bohr magnetons. It is usual to assume for this purpose that the 4f electrons obey Russell–Saunders coupling, and that the ground state is given by Hund's rule (Van Vleck 1932, p. 241). With the exception of samarium, the experimental results show that above a certain temperature the susceptibilities of the metals obey a Curie–Weiss law $\chi = C/(T-\theta)$ (or a modified Curie–Weiss law $\chi = C/(T-\theta)+K$ in the case of neodymium). In the temperature range where this type of law is followed we may assume that the Curie–Weiss θ may be ascribed to interactions between neighbouring ions, so that μ_{eff} may be calculated from the equation $\mu_{\text{eff}}^2 = 8{\cdot}05CM$ where M is the atomic weight of the metal. The theoretical and experimental values of μ_{eff} are compared in the table, and the comparatively good agreement for Ce, Pr and Nd shows that in these metals the assumptions of a nearly 'free' 4f sub-shell is justifiable. The case of samarium is complicated by the fact that in this element the L, S coupling is rather weak, and the redistribution of the ions between the ground level ($j=5/2$) and the next level ($j=7/2$) becomes even more important than in the case of Nd (Van Vleck 1932, p. 245ff). If in addition we have to take into account neighbouring-ion interactions the situation becomes complicated, but a very rough comparison between theory and experiment can be made by plotting the theoretical relation between χ and T given by Van Vleck (1932, p. 251) and assuming that the effect of the interactions is to shift the origin of temperature from $T=0$ to $T=-\theta$ (i.e. the whole curve is moved to the left a distance θ). Figure 6 shows this plot (broken line) assuming $\theta = 60$°K, and it is seen that in order of magnitude the calculated susceptibility agrees with experiment.

The three rare-earth metals other than praseodymium have a maximum susceptibility at the Néel temperature T_N°K shown in the table. $\chi_{\max}$, the value of the susceptibility per gramme atom at this temperature, is also given, that for praseodymium being the temperature-independent value below about 4°K. Specific heat measurements (Parkinson, Simon and Spedding 1951, Parkinson and Roberts 1957, Roberts 1957) show marked peaks in the atomic heats of cerium, neodymium and samarium at temperatures very close to those at which the susceptibility reaches its maximum (see table), and the shapes of these peaks, falling off rather steeply on the high-temperature side, suggest some kind of co-operative phenomena. Together with the maxima in the susceptibility the evidence suggests that in these metals anti-ferromagnetism sets in at the temperature T_N, an occurrence whose possibility in the case of the rare earths has already been suggested by Tredgold (1954). In the case of praseodymium the atomic-heat curve (Parkinson, Simon and Spedding 1951) exhibits a broad maximum only, and appears to be fully accounted for by the redistribution of the 4f electrons among the different levels arising from Stark splitting of the ground state $^3H_4$. The constancy of the susceptibility of praseodymium at low temperatures indicates that the lowest of these levels must be a singlet. It is probable that for anti-ferromagnetism to occur the lowest magnetic level must be at least doubly degenerate, and the theorem of Kramers (1930) states that this is necessarily the case only for magnetic ions with an odd number of electrons, as in cerium, neodymium and samarium. Hence the appearance of anti-ferromagnetism in these three metals and its absence in praseodymium may well be a consequence of the former having odd numbers of 4f electrons and the latter an even number.

Element	Term	μ_{eff} (μ_B)		T_N(°K)		$10^3\chi_{max}$/g atom
		Theory	Expt.	Max. Susc.	Sp. Ht.	
Ce	$^2F_{5/2}$	2·54	2·51	12·5	12·5	16·1
Pr	$^3H_4$	3·58	3·56	—	—	217
Nd	$^4I_{9/2}$	3·62	3·3	7·5	7·2	240
Sm	$^6H_{5/2}$	†	†	14·8	13·6	1·82

† Temperature dependent.

The case of cerium is complicated by the thermal hysteresis phenomenon occurring between about 40°K and 200°K. This has been ascribed to the transfer of some of the available 4f electrons to a 5d state (Shuch and Sturdivant 1950, Parkinson, Simon and Spedding 1951), a suggestion which explains the drop in susceptibility which occurs on cooling a virgin specimen just below 100°K. The gradual shrinkage of the hysteresis loop on successive coolings and warmings may possibly be due to the appearance of a small proportion of hexagonal close-packed structure in the face-centred cubic material (see above). This might well induce strains large enough to modify the relative energy levels involved and inhibit the 4f → 5d transition.

Whatever the cause, the susceptibility at and below the Néel temperature T_N shows that the number of 4f electrons per atom is considerably lower for a specimen cooled for the first time than for a specimen in the almost reversible condition produced by 100 coolings and warmings. Assuming that for the latter the number is almost 1 per atom, the maximum susceptibility of the virgin specimen $8{\cdot}2 \times 10^{-3}$ e.m.u. per gramme atom compared with $16{\cdot}1 \times 10^{-3}$ e.m.u. per gramme atom for the 'reversible' material, suggests that about half of the available 4f electrons go into the 5d state in the initially cooled cerium. A rough confirmation of this comes from careful specific heat measurements (Parkinson and Roberts 1957) on a cerium specimen from the same material as that from which one of our specimens was prepared. They find that the peak in the atomic heat at 12·5°K gradually increases in size when a virgin specimen is subjected to the same successive cooling and warming treatment as our specimens, the final peak containing about double the entropy of the initial peak.

In the case of neodymium the change in slope at 20°K of the $(1/\chi, T)$ plot is accompanied by a second peak in the specific heat (Parkinson, Simon and Spedding 1951). This remains at present unexplained—it seems unlikely that redistribution of the 4f electrons between a basic doublet (anti-ferromagnetic) state and higher levels could cause either such a sharp change in the gradient of the susceptibility curve, or such a steep drop in the atomic heat just above 20°K.

The susceptibility of both cerium and samarium begins to rise again at the lowest temperatures attained, and in the case of cerium this rise is noticeably more marked in the initial cooling. It is perhaps due to the 5d electrons in the metal—lanthanum, too, shows a rise in susceptibility as the temperature is reduced (see figure 1). For praseodymium and neodymium the susceptibility of the 4f electrons is so great that the small effect of the 5d electrons would not be noticeable.

§ 5. Conclusion

Pure elements exhibiting anti-ferromagnetism are not common. The evidence put forward here for its occurrence in certain of the rare-earth metals ought therefore to be tested by other means. An unambiguous proof of anti-ferromagnetism would probably be afforded by neutron diffraction experiments at low temperatures.

Acknowledgments

The author wishes to thank Dr. D. H. Parkinson and Mrs. L. M. Roberts for helpful discussions at all stages of this work, and Mr. A. G. Harris for his valuable assistance in the experimental measurements. He also wishes to thank Miss A. E. Rennie for the x-ray photographs and analysis. This paper is published by permission of the Controller, H.M. Stationery Office.

References

Bates, L. F., Leach, S. J., Loasby, R. G., and Stevens, K. W. H., 1955, *Proc. Phys. Soc.* B, **68**, 181.

Elliott, J. F., Legvold, S., and Spedding, F. H., 1954, *Phys. Rev.*, **94**, 50.

de Haas, W. J., and Gorter, J. G., 1930, *Commun. Kamerlingh Onnes Lab. Leiden*, **208** c.

Kramers, H. A., 1930, *Proc. Acad. Sci. Amst.*, **33**, 959.

La Blanchetais, C. H., 1945, *C.R. Acad. Sci., Paris*, **220**, 392.

Parkinson, D. H., and Roberts, L. M., 1957, *Proc. Phys. Soc.* B, **70**, 471.

Parkinson, D. H., Simon, F. E., and Spedding, F. H., 1951, *Proc. Roy. Soc.* A, **207**, 137.

Roberts, L. M., 1957, *Proc. Phys. Soc.* B, **70**, 434.

Shuch, A. F., and Sturdivant, J. H., 1950, *J. Chem. Phys.*, **18**, 142.

Tredgold, R. H., 1954, *Proc. Phys. Soc.* A, **67**, 221.

Trombe, F., 1934, *C.R. Acad. Sci., Paris*, **198**, 1591.

Van Vleck, J. H., 1932, *The Theory of Electric and Magnetic Susceptibilities* (Oxford: Clarendon Press).

Yost, D. M., Russell, H., and Garner, C. W., 1947, *The Rare-Earth Elements and their Compounds* (New York: John Wiley), p. 63.

The Statistical Mechanics of the Steady State

By H. N. V. TEMPERLEY
Atomic Weapons Research Establishment, Aldermaston, Berks.

MS. received 23rd May 1956, and in revised form 3rd December 1956

Abstract. It is shown that the Fowler–Darwin definitions of the temperature and chemical potential can be extended in a natural way to cover non-equilibrium but steady assemblies. The concept of equilibrium is replaced by the more general one of 'compatibility', and it becomes possible to show that the proposed statistical definitions of quantities like temperature reduce to suitable averages over the assembly of the locally measured quantities. The usual restrictions to small departures from equilibrium are unnecessary. A simple problem (flow of gas in a tube) is discussed from the new point of view.

§ 1. INTRODUCTION

THE present paper arose from an attempt to translate into statistical-mechanical language the well-known 'textbook' treatment of the Joule–Thomson effect, based on elementary thermodynamics. In this paper, we shall suggest that a statistical mechanical treatment of a steady assembly, such as a gas or liquid flowing through a pipe, is possible. The thermodynamic treatment of a steady-flow problem is possible because macroscopic conditions at any one point are effectively constant in time. A very large class of apparently non-steady phenomena can be made macroscopically steady by a proper choice of the boundary conditions, for example a shock front can be held in a tube at a constant position.

We shall conclude that the Fowler–Darwin formalism of statistical mechanics *can* be extended in a natural way to cover many, if not all, steady assemblies, even if they are a long way from static equilibrium. The starting point of any statistical-mechanical treatment must be some set of energy levels (discrete or continuous), which often refer to single particles, but must, in a general assembly with interactions, involve *all* the particles. Now it is, in the first place, clear that non-trivial steady assemblies do exist, which can be handled by Schrödinger's time-independent wave equation, such as the impact of a stream of particles on a potential barrier, or the flow of a stream of particles through a tube of varying cross section, or the deflection of a stream by a fixed scattering centre. For non-interacting particles, we are free to work with single-particle levels, and it does not matter whether these are sharp or diffuse, discrete or continuous because the sum-over-states goes over into the appropriate integral. Equally, we could work with the energy levels associated with all the particles that are, at any instant, within a certain region of space such as a tube. The solution of an ordinary scattering problem does not have to take account of the means by which the particles are accelerated and detected, since it is always possible to assume that the relevant devices are a long way from the part of the assembly we are studying. With, for example, particles driven steadily through a tube by a piston, the distance of the piston from the region we are studying hardly matters at all. If we have a tube of varying cross section, we show in §7 that a satisfactory account of the Bernoulli variation of pressure can be given for a perfect gas, using the definition of local pressure given in equation (4) below, which has thus been checked in a simple case.

A more difficult point arises when we consider assemblies of *interacting* particles. One might think that the only effect of the interactions, if weak, is to modify the energy-levels, and if strong, is to make it necessary to use many-particle rather than single-particle energy levels. In this paper, we shall assume that this is so, relying on the analogy with classical assemblies, for which normal modes always exist, and we shall examine some consequences of this assumption. It is, at present, impossible to check it theoretically, because no realistic calculation of, say, a stream of interacting particles incident on a barrier, or travelling through a converging tube, ever seems to have been made quantum-mechanically, while, if we considered a Hooke's law solid, for which exact quantum-mechanical calculations *can* be made, this has the serious disadvantage that it shows no dissipative effects, and therefore, like the perfect gas, *may* be a very special case. Meanwhile, it seems worth examining the consequences of the above assumption. An obvious way in which an interacting assembly differs from a non-interacting one is in the existence of the Joule–Thomson effect with which we began—an imperfect gas flowing in a converging tube shows differences in local temperature. We shall show that, on our assumption that energy levels can still be defined, these variations of local temperature can be described in a natural way. Even should it turn out that the concept of energy levels is formally invalid, the proposed treatment may still be a reasonable approximation for assemblies with small interactions or for assemblies of the type described in §7. Again, occurrences such as the effective broadening of some of the levels would make no real difference to the treatment.

With this proviso, we shall arrive at natural, and mutually consistent, definitions of the 'local' temperature of a small region and of the temperature of the whole assembly, similar remarks apply to quantities like pressure. These definitions are based solely on statistical mechanics, and make no explicit reference to heat engines, but become identical with the accepted definitions when applied to an assembly in equilibrium. There is no question of a *difference* between 'statistical' and 'thermodynamic' temperatures, since we do not know how to define the latter for a non-equilibrium assembly. What we *do* have to recognize is that an assembly in a steady state requires *more* macroscopic parameters to describe it than does a similar assembly in equilibrium.

§ 2. THE TYPE OF ASSEMBLY STUDIED

The assembly that we study consists of a gas, or isotropic liquid, moving steadily through a tube of non-uniform cross section. It must incorporate some means of measuring local temperatures and pressures, and we shall describe, in a moment, an idealized mechanism that is capable of acting, at will, as a pump, a manometer or thermometer or as a means of introducing or withdrawing heat at any point, and we shall suppose that our Schrödinger equation describes the molecules in a given length of tube, together with these auxiliary devices that are coupled to it. Our idealized 'measuring instrument' is extremely simple, yet is capable of measuring both pressure and temperature. We may suppose that small side tubes are let into our tube at various points along its length, each of them closed by a piston backed by a spring. First of all, let us suppose that the only coupling between one such device and the rest of the assembly is through the collision of molecules with a piston. The spring will be compressed to an average extent that measures the ' pressure ' in the assembly at the point. Furthermore, it will oscillate about this mean position as the result of the collisions, and the mean energy associated with these oscillations, the Brownian movement, is a measure of the ' temperature ' of the assembly at that point, which can be measured in principle if we measure the mean square displacement of the piston over a reasonable number of periods. Now let us suppose that we increase the amplitude of the oscillations of the piston by supplying energy to it mechanically or electrically. This extra energy will be imparted to the molecules colliding with the piston, which thus becomes a source of heat. We could equally well make a heat ' sink ' by connecting the piston to a dashpot or to an electromagnetic brake, that is, by coupling it in some way to some assembly of lower temperature than the one we are studying. In the present paper we shall be concerned simply with these pistons as manometers and thermometers.

§ 3. THE CONCEPT OF ' COMPATIBILITY ' OF STEADY-STATE ASSEMBLIES

The reasoning of thermodynamics and statistical mechanics is based very largely on the concept of equilibrium between two different assemblies, that is, nothing happens to the macroscopic properties of two such assemblies if we permit exchanges of energy or matter, or both, between them. In other words, in certain circumstances we can bring them into intimate contact, or even incorporate them with one another, without any changes in pressure, temperature etc. Now it is physically obvious that pairs of steady-state assemblies also exist with an analogous property, and it is this property that we define as *compatibility*. Like all new concepts, it is best understood by means of examples.

(*a*) Two streams of gas, flowing in two tubes of a heat-exchanger, will be compatible and no macroscopic exchange of heat will occur if their temperatures are the same all along the dividing wall. If, further, their pressures and densities are also the same, and the streams are chemically alike, the streams are compatible and will not affect one another macroscopically even if the dividing wall is *porous*. Finally, if their tangential velocities are made to agree, the two streams are compatible even if there is no dividing wall at all.

(*b*) Two samples of gas, with pressures, temperatures and stream velocities corresponding to two stages in the history of a sample of gas undergoing a Joule–Thomson expansion or some other steady-state process, would be said to be compatible. This is because either or both samples could be incorporated into a certain steady assembly without any change in the observable properties, provided that we chose the appropriate region of that assembly for each sample. In each case we could add the sample to the assembly, provided that the temperature, pressure, stream velocity, etc., of the small sample agreed with the *locally measured* values of these quantities for the assembly.

(*c*) Consider an assembly such as a gas conducting heat. We might divide it up into small samples by planes along and perpendicular to the heat flow. We could imagine the assembly built up of these small samples which could, in principle, be added one by one without disturbing the temperatures or gradients in the rest of the assembly. Although two such neighbouring samples may have slightly different local properties, they are compatible. Some of their properties must agree, for example the temperature and heat flux along their common boundary, but others, such as their mean measured temperatures, will be different. Yet, if they are brought together in the proper way, no changes in either set of properties take place.

Just as we begin thermodynamics with the argument that two assemblies in equilibrium must have some properties in common, and then arrive at definitions of these properties from a study of the conditions of equilibrium, so we now attempt a similar programme for compatible assemblies. We shall show that it is possible to define a temperature and entropy for the whole of a steady assembly, and that these quantities correspond to appropriate averages over the locally measured values of these quantities. The 'local' measurements could consist of observations on pistons in side tubes, or the use of more practical thermometers and manometers.

§ 4. The Statistical Mechanics of the Steady State

In equilibrium, we work with two main types of ideal assembly, the completely closed assembly whose total energy and total number of molecules are held fixed, and the open assembly which is permitted to exchange energy and matter with heat-baths and reservoirs. The open assembly corresponds to the case in which we fix our attention on a particular length of tube, through which gas is being driven steadily. The energy and number of molecules in the assembly fluctuate about mean values as the result of interchanges with neighbouring portions of the gas which are compatible with it. (The analogue of the 'closed' equilibrium assembly will not be considered in the present paper. We could arrange that the energy and number of molecules remain constant, but we could not close it off completely from its surroundings on account of the necessity of supplying energy to maintain it in steady motion and of withdrawing a corresponding quantity of heat. This means that the assembly would have to be described by a *time-dependent* Schrödinger equation, but this does *not*, by itself, invalidate the assumption that it possesses approximately stationary energy levels. We give below (§ 7) some examples of non-trivial assemblies for which such an assumption seems very plausible.)

We therefore make an assumption analogous to that in the equilibrium case, namely that our assembly, when it contains N molecules, has $g(N, E_r)$ energy levels, each of *total* energy E_r, so that we do not make the usual restriction to small interactions, but imagine that we have solved the many-body wave equation. We also, as in the equilibrium case, need consideration of the way in which these numbers of levels vary with N in order to be able to handle the 'open' assembly. We also make the convention that the E_r's are to include the entire kinetic energy, including that of any bulk motion, of the gas or liquid in the assembly, but that we specify separately the kinetic and potential energies of the small pistons and springs that we are coupling into the assembly as measuring instruments. We are keeping the discussion quite general, so that we are assuming implicitly that we only list energy levels corresponding to wave functions with the correct symmetry properties and that any factors resulting from spin degeneracy are incorporated into $g(N, E_r)$.

Now, just as in equilibrium theory, we can, on the assumption that energy levels exist in some sense, define the grand partition function by the equation:

$$f(\lambda, \theta) = \sum_N \sum_{E_r} \lambda^N \theta^{E_r} g(N, E_r) \qquad \ldots\ldots(1)$$

which equation can nearly always be asymptotically inverted to read

$$\log g(N, E) \simeq \log f - N \log \lambda - E \log \theta$$

with

$$\lambda \frac{\partial}{\partial \lambda}(\log f) = N \text{ and } \theta \frac{\partial}{\partial \theta}(\log f) = E. \qquad \ldots\ldots(2)$$

Now, a completely isolated assembly for which N and E are fixed is, at any given instant, equally likely to be found in any one of the $g(N, E)$ complexions of given N and E. We have, in equilibrium theory, to discuss the interaction of at least two assemblies in order to be able to define the concept of temperature. The argument runs as follows: In an actual assembly, it turns out that the main contribution to (1) comes from very small ranges of values of N and E, the quantity written $g(N, E)$ in (2) representing, strictly speaking, a summation, or integration, over these critical ranges, rather than one single term of (1). This is best established by the Fowler–Darwin process of contour-integration, which can then be used to justify, and replace, the older, and more intuitive, principle of 'maximum thermodynamic probability'. If now we consider two assemblies, they can, in principle, be coupled together in such a way that energy or matter can pass between them, but the opening of a small passage need only cause negligible changes in the energy levels. In these circumstances, the two assemblies together can be treated as one; the energies are additive, so that the partition functions (1) can be multiplied and the Fowler–Darwin process can again be used to locate values of λ and θ. If the assemblies were in equilibrium no changes in these quantities occur, if not, they adjust themselves to values intermediate between those for the two separate assemblies, and the 'steepest descent' argument then shows that the available matter and energy distribute themselves between the two assemblies in such a way that $g_1(N_1, E_1) \times g_2(N_2, E_2)$ is a maximum, where the g's are to be understood as mean values in the sense of equation (2) rather than as single coefficients as in equation (1).

If we can define expression (1) for steady assemblies, the mean N's and E's will not change when two of them are coupled together, if they are compatible. The mathematical argument proceeds as before. If they are not compatible, changes in N and E will take place, and this will be associated with an increase in the product of the g's and an equalization of the λ's and θ's. The mathematical formalism is completely unchanged, and we can therefore take the whole of Fowler's argument (1936) to show that two assemblies which are allowed to share energies are described by the same θ in their partition functions, and that, if they are allowed to share matter, they are described by the same λ. (We cannot yet identify these quantities with temperatures or partial potentials because the pressures and temperatures vary from point to point in a steady assembly.) For the moment, all that we assert is that these quantities each represent *some common property of two compatible assemblies*, two assemblies being defined as compatible if there exists a third steady assembly into which both of the assemblies considered can be incorporated without any changes in their macroscopically measurable properties.

We now show that, if an assembly is described by a partition function such as (1), we can arrive at a satisfactory definition of the local pressure at each point. Consider what happens to the assembly if we alter, by a small distance dx, the position of the piston in one of the side tubes, which tube we suppose to communicate with the assembly at a distance y along the main tube. The result of the displacement is a slight change in the volume of the assembly and a corresponding shifting of all the energy levels. If we knew for certain that the assembly was occupying a certain level r of the energy E_r, the change in the energy of that level would exactly balance the work needed to compress the the spring through dx, so that we should have $P_1 = -(1/a)(\partial E_r/\partial x_1)$, where a is the cross section of the side tube. Actually, the assembly is almost equally likely to be found in a very large number of different levels of nearly the same energy, so that we have

$$P_1 = -\frac{1}{a}\left[\frac{\partial E_r}{\partial x_1}\right]_{\text{av}} \qquad \ldots\ldots(3)$$

where the averaging has, just as in the equilibrium case, to be performed on the basis that a single level of total energy E_r is occupied by a total of N molecules with a probability proportional to $\theta^{E_r}\lambda^N$, that is

$$\begin{aligned} P_1 &= -\frac{1}{a}\sum_N\sum_r\left(\frac{\partial E_r}{\partial x_1}\right)\theta^{E_r}\lambda^N \Big/ \sum_N\sum_r \theta^{E_r}\lambda^N \qquad \ldots\ldots(4) \\ &= -\frac{1}{a\log\theta}\frac{\partial(\log f)}{\partial x_1} \end{aligned}$$

a similar formula applying to all points of the assembly. This is a natural generalization of the formula $P = -(1/\log\theta)[\partial(\log f)/\partial V]$ applying to an equilibrium assembly (in which the pressure is uniform, so that it does not matter in what way the deformation of volume is supposed to take place). Thus, in the steady assembly, the single pair of ' force ' and ' displacement ' variables (P, V) characteristic of an equilibrium assembly is replaced by a whole set of pairs of such variables, one pair for each point along the tube containing the assembly. In what follows, such ' local ' variables will be noted by a suffix, or will be expressed as functions of y, e.g. P_1, $P(y)$.

We may also interpret equation (4) slightly differently. It describes what happens if a small volume adx_1 is removed from the side tube and incorporated with the main assembly, the result being that the quantity $-\log f/\log\theta$ is increased by P_1adx_1. Now we can always, in principle, consider an assembly as being built up, step by step, out of small elements, provided that this can be done in such a way as to avoid changing the measured properties of the part of the assembly that is already there, that is, provided that the portions added at each stage are *compatible* with the rest of the assembly. Thus, we might build up our assembly, consisting of gas flowing through a tube, by injecting at each point gas at the appropriate pressure, velocity and density, the walls of the vessel expanding sideways at the appropriate rate at each point. If this were done properly, all parts of the assembly could be described by the same values of θ and λ throughout the process. Each increase in the cross section of the assembly at any point would result in a change in all the energy-levels, so that the values of these would determine the value of $\log f/\log\theta$ at each stage. (We have restricted ourselves to cases where the pressure is known to be isotropic as a matter of observational fact. Generalization of this work to the case where the stresses have to be represented by a tensor is almost certainly possible.)

On this basis we conclude that $-\log f/\log\theta$ is an additive quantity for compatible assemblies, in other words that

$$-\frac{\log f}{\log\theta} = \int P(y)A(y)dy. \qquad \ldots\ldots(5)$$

Massignon (1955), obtains a very similar result for equilibrium, but non-uniform assemblies (see his equation (4)). In equation (5) the integration is to be taken over the assembly, $P(y)$ being the locally measured pressure at a point y along the tube and $A(y)$ the corresponding cross section. This is to be compared with the result obtained by Fowler (1938) for an equilibrium assembly

$$-\frac{\log f}{\log \theta}=\sum_{s} X_{s} x_{s} \qquad \ldots\ldots(6)$$

where the summation is over all relevant pairs of ' force ' and ' displacement ' variables, which sum reduces simply to PV if pressure is the only force variable. Comparison of equations (5) and (6) suggests that we can define a ' statistical temperature ', by means of the variable θ, for a steady assembly as well as for an equilibrium one. We can also define, if we wish, a mean pressure $\bar{P}$ for the whole assembly by writing

$$\bar{P}V=-\log f/\log \theta \qquad \ldots\ldots(7)$$

and equation (5) then shows that $\bar{P}$ represents the measured pressures averaged in a natural way over the whole assembly.

Other standard formulae can be generalized in a natural way. Once it is realized that $\log f$ is additive for two compatible assemblies we can apply the equations $E=\theta\partial(\log f)/\partial\theta$ and $N=\lambda\partial(\log f)/\partial\lambda$ to a portion of our assembly quite as well as to the whole of it, the function replacing $\log f$ in these formulae being obtained by restricting the integration in equation (5) to part of the assembly, whereupon the above two formulae lead to the common sense result that, if we want the mean number of molecules likely to be found in the assembly between y and $y+dy$ we must write

$$N(y)dy=\lambda\frac{\partial}{\partial\lambda}\left(\partial\frac{(\log f)}{\partial y}\right)dy \qquad \ldots\ldots(8)$$

and similarly

$$E(y)dy=\theta\frac{\partial}{\partial\theta}\left(\partial\frac{(\log f)}{\partial y}\right)dy \qquad \ldots\ldots(9)$$

for the total kinetic and potential energy residing between these two planes.

§ 5. THE CONCEPTS OF ' LOCAL ' AND ' STATISTICAL ' TEMPERATURES

These considerations, together with the equations (5)-(7), lead us to look for a connection between the statistical parameter θ and the measured *local* temperatures, the additivity of $\log f$ suggesting the existence of a result similar to that found from (5) and (7) for the pressure.

We can deduce the form of the expression for local temperature by a somewhat different argument, and we shall arrive independently at the expected relation between θ and an average over the local temperatures. Consider equations (3) and (4) for the local pressure. This derivation is based on the argument that, if our assembly were definitely known to be in energy level E_r, there would be a definitely known mean compression of the spring of the piston at station 1, and at all other stations where we had inserted side tubes. This follows directly from the fact that there is supposed to be a definite coupling between the pistons and the rest of the assembly, which are described by the same many-body Schrödinger equation.

The *mean square* displacement of piston 1 is also fixed if the assembly is definitely known to be occupying the level E_r. If the assembly is occupying other levels in turn, each level is associated with a certain mean square displacement of piston 1 and the mean square displacement that we shall actually observe will again be an appropriate average over all the levels of the assembly, of the same type as (3) and (4). In the language of statistical mechanics, we say that piston 1 is trembling about its mean displacement because it is coupled with the rest of the assembly and that the amplitude of this trembling is not quite constant in time because the assembly is continually going into new energy levels. In the language of kinetic theory, we say that this trembling is caused by bombardment of the piston by neighbouring molecules, and that the mean square amplitude is a measure of the local ' temperature '. Clearly there is no reason to expect that the ' readings ' shown by two pistons in different parts of the assembly should be the same, any more than there is reason to expect *pressures* to be uniform in a non-equilibrium assembly. On the other hand, a quantity such as θ, determining the probabilities of occupation of all the levels, is a property of the assembly as a whole, while the ' local ' temperatures refer to particular places, but are statistical averages over all the accessible levels. Our central problem is to relate these two quite distinct definitions of temperature, but before we can do this, we must determine the process of averaging over the levels by which ' local ' temperatures are defined, analogous to equation (3) for the local pressure.

We can determine this quantity by analogy with (3). The quantity $dx_1(\partial E_r/\partial x_1)$ may be thought of as the change in the energy level E_r that would be brought about by applying an external force to piston 1 which changes its equilibrium position by dx_1. Equally, we could think of it as the change in E_r that results from injecting an additional volume adx_1 of *material* into the side tube. (This extra material would have to be compatible with that already present, which implies that its pressure and temperature would have to be P_1 and T_1). We can conclude that these two slightly different operations must have almost the same effect on the energy levels from the undoubted experimental fact that spring or Bourdon, and liquid or U-tube manometers give consistent results. We now examine instead the effect of introducing a small additional quantity of energy into the assembly at station 1. This might be done in a number of ways, but we shall suppose that we bring up to the side tube 1 a body at a temperature infinitesimally greater than T_1 until the energy $v_1T_1ds_1$ is added to the assembly. Here v_1 is the volume of the side tube and s_1 the entropy content per unit volume of the material of the assembly in equilibrium at P_1 and T_1. We must now examine the effect on the rest of the assembly of this slight change in s_1 which, of course, in general implies a corresponding change in T_1.

If the assembly were definitely known to be in the level E_r, this slight change in one of the local temperatures would mean simply that the energy content of the assembly would be slightly greater, and we should have $dE_r = v_1T_1ds_1$, this being a simple consequence of the conservation of energy. What we are really asserting is this; to every given energy level of the whole assembly there would, as we have pointed out already, correspond a given set of amplitudes of oscillation of the pistons, and therefore a unique set of local temperatures. If we alter one or more of these temperatures, our assembly is subject to a slightly different set of constraints, and it is the corresponding effect on the energy levels that we are trying to determine. It would be more natural to define our changes of energy level in terms of changes of these amplitudes, that is, in terms of changes in quantities like T_1; the use of the local entropy per unit volume instead is purely formal, and is adopted in order to get a closer analogy with formulae (3) and (4) above.

To take account of the fact that a large number of different energy levels are consistent with the same boundary conditions, we postulate a formula of the type

$$T_1 = \frac{1}{v_1}\left[\frac{\partial E_r}{\partial s_1}\right]_{\text{av}} = \frac{\sum\limits_N \sum\limits_r (\partial E_r/\partial s_1)\theta^{E_r}\lambda^N}{v_1 \sum\limits_N \sum\limits_r \theta^{E_r}\lambda^N} \qquad \ldots\ldots(10)$$

to allow for the fact that the assembly may be found in a large number of different levels. We may not write this relationship simply as

$$T_1 = (1/v_1 \log\theta)\,[\partial(\log f)/\partial s_1],$$

because the introduction of a small quantity of energy into the assembly not only changes the levels E_r, but also, in general, changes the probability of occupation of each level, that is, we must allow for the effect of small consequential changes in θ and λ. From (2) it is plain that (10) is equivalent to

$$T_1 = \frac{1}{v_1 \log\theta}\frac{\partial}{\partial s_1}(\log f - N\log\lambda - E\log\theta) \qquad \ldots\ldots(11)$$

because then the coefficients of $\partial\lambda/\partial s_1$ and $\partial\theta/\partial s_1$ vanish by (2).

If the assembly is such that the local entropy and temperature are functions of y only, that is, if we may consider them to be the same over any cross section perpendicular to y (we shall return to this point later), we may arrive at a formula analogous to (5) by integrating (11) over the entire assembly, the entropy content of the region between y and $y+dy$ being simply $A(y)S(y)dy$ where $S(y)$ is the entropy content per unit volume at y.
Then

$$E + \frac{N\log\lambda}{\log\theta} - \frac{\log f}{\log\theta} = TS = \int A(y)T(y)S(y)dy. \qquad \ldots\ldots(12)$$

We can carry through an exactly similar discussion for the other selector variable λ by imagining a process in which extra molecules are added to the assembly at station 1, in such a way that the statistical temperature θ is unaltered. By considering the quantity $[\partial E./\partial N_1]$, we arrive at the result for the Helmholtz free energy in a very similar fashion :

$$F = T\int \frac{\mu(y)A(y)\rho(y)}{mT(y)}\,dy \qquad \ldots\ldots(13)$$

where $\mu(y)$ is the measured value of the chemical potential at station y and $\rho(y)$ is the local density, and m is the mass of one molecule.

We must now consider how to define the local entropy content per unit volume s_1. Our extension of statistical mechanics calls for a close relationship between this quantity, the local temperature T_1 and the energy changes caused by feeding a given quantity of heat into the assembly along side tube 1, or by feeding matter into the side tube, which necessarily carries energy with it, but we have still to show that we can consistently give a definition equivalent to the 'equilibrium' one in terms of specific heats. Now we defined T_1 in terms of the Brownian motion of a piston in side tube 1, but we could equally well have defined it by saying that a further small assembly could be used as a local '*thermometer*'. That is to say that, if it were initially in equilibrium at the temperature T_1 it could be brought into thermal contact with side tube 1, without suffering any net gain or loss of heat. That is, it would be compatible with the main assembly. Now suppose that, instead of the thermometer, we introduce into the assembly along side tube 1 a small mass of material that is initially in equilibrium at the local temperature and density. Again, no net interchange of energy or matter with the rest of the assembly will occur, so that it is plain that the energy so added to the assembly is determined by the *equilibrium* heat capacity of the small additional mass, that is that we can safely define entropy content in the side tubes in terms of *equilibrium* heat capacities. Our previous work then shows us that, if we extend our definition of local entropy from each side tube to regions of the main assembly in the same plane, we remain consistent with the statistical definition of the entropy of the whole assembly, the possibility of adding up entropies being equivalent to compatibility.

In making our definition of compatibility, it was understood that we should have to make restrictions on the possible ways in which two assemblies are going to be brought into contact. There is, therefore, no *general* principle that two assemblies that are compatible with a third assembly are compatible with one another, though more restricted statements of this type are possible in special cases. We find, for example, that there is a distinct difference between the concepts of compatibility between two assemblies that are to have only one very small region in common, and between two assemblies that are to make contact at two or more widely separated points or over a region of finite size, such as the common boundary along the direction of heat flow. Let us suppose that one of the steady-state assemblies is a non-equilibrium one, so that its local properties differ from point to point. Then, as we have seen, it is possible for a small assembly, originally in equilibrium, to be compatible with this one if it is brought into contact with it over one small region only, but if contact occurs over a region large enough for a significant change in the local properties to occur, then compatibility is possible only if the second assembly is also a non-equilibrium one. Bound up with this is the fact that, although there is always a sense in which we can say that two spatially separated small portions of a steady assembly are described by the same θ, it does *not* follow from this that they have the same *local* temperature. (None of the above work has envisaged such a process as bringing two spatially separated parts of a non-equilibrium assembly into contact, although this might be possible for an equilibrium one. A finite deformation of any assembly implies that all its energy levels are altered, and in general, this would change the relationship between θ and E. Again, if we attempted to separate off one small portion of an assembly in order to bring it near another portion, it would, once separated off, have to be described by a set of wave functions quite different from those describing the main assembly.)

It is a trivial point that we cannot even define equilibrium between two assemblies apart from some expressed or implied specification of the means by which they are to be coupled up (e.g. thermal contact, semi-permeable membranes). Two assemblies could be in equilibrium if in thermal contact, but not in equilibrium if transfer of matter was allowed. The restrictions of the definition of compatibility are of the same type. They are, in practice, more complicated simply because we require a more elaborate macroscopic description of a steady-state assembly, more parameters being involved than for an equilibrium one.

§ 6. Discussion

We have shown that, for a steady assembly, we can generalize the concepts of ' statistical temperature ' and ' statistical partial potential ', arrived at by an extension of the Fowler–Darwin treatment of the equilibrium assembly, and that these quantities can be expressed as appropriate averages over the *locally measured* properties of the assembly. There is one important proviso, for a steady streaming assembly such as a gas flowing through a tube. The measurements are supposed to take place in side tubes, which means that both the instruments and the material in immediate contact with them are at rest. In practice, measurements of pressure and temperature are nearly always carried out in this way, but there are also types of measurement, such as the use of a Rayleigh disc, that depend essentially on the local stream velocity. We have not considered the effect of this explicity, what we are saying is that, even in a portion of the assembly where the stream velocity is appreciable, we can consistently *define* such quantities as local temperature and entropy by reference to a neighbouring portion of the assembly that is at rest, but is compatible with the part that we are talking about. We assume always that conditions have settled down so that these two portions do not, in the long run, exchange matter or energy with one another even though they may be free to do so. Since there is, at present, no precise thermodynamic definition of temperature or entropy in part of an assembly in macroscopic motion, and since measurements are seldom, if ever made with instruments ' moving with the stream ', it seems that there is no inconsistency in such an extension of our definitions, which is, moreover, in keeping with the accepted ways of measuring temperatures in moving fluids.

Some recent work by Massignon (1955), appears to be closely related to the first three sections of this paper. He sets up definitions of thermodynamic quantities in terms of average fluctuations. It is well known that such definitions are possible in the equilibrium case, but he shows that they can be applied to define *local* thermodynamic properties (in e.g. an assembly exposed to an external field of force), and his claim that such a *non-uniform* assembly can be divided up into portions each making its own contribution to the free energy appears to be closely related to ours. Compare his equation (4) with our equation (5). So far, he has only dealt with equilibrium assemblies, but, since he makes use of the concept of a quasi-static change, his work should be capable of further applications, for example to the cases of non-zero momentum that we deal with below (§ 7).

The contention that an assembly has stationary states when subjected to a non-uniform external force is certainly correct; what is more doubtful is whether the same is true if the *temperatures* are also constrained to be non-uniform. If this is so, then the formal apparatus of the remainder of the present paper can be used to describe the responses of the steady assembly to changes in these constraints. There seems no real difficulty in thinking of an assembly whose motion is maintained, say by absorbing radiation from its surroundings and emitting an equivalent amount. Indeed, the ' Electrolux ' type of refrigerator, or an ordinary central heating system, could be made to approximate to this situation fairly closely, and the assumption that such assemblies have large numbers of approximately stationary states that they occupy in turn, seems eminently reasonable, though we do not at present know how to handle the quantum mechanical problems.

§ 7. Treatment of a Simple Problem

We consider a perfect gas flowing in a rectangular tube of cross section $Q \times R$, with a steady stream-velocity u. We shall examine how the pressure varies with u if the density and temperature are kept constant. To calculate the local pressure, we shall use formula (4), by fixing our attention on a short length of tube L, and examine the effect of varying the cross section slightly by pushing one of the walls inwards. The wave functions appropriate to this problem are

$$\psi = e^{\pm ipx} \sin\left(\frac{q\pi y}{Q}\right) \sin\left(\frac{r\pi z}{R}\right) \qquad \ldots\ldots(14)$$

where q and r are integers, but p may take any real value. We have now to consider how we may allocate our occupation numbers between states $+p$ and $-p$ to take account of the fact that the mean stream velocity of the whole assembly is known to be u, that is to say that the total x-momentum in the given portion of tube is $nLQRmu$ where $nLQR = N$ is the number of molecules in this portion. This problem was considered by Temperley (1952 Appendix) and by Zilsel (1953). Our occupation numbers now have to be chosen in such a way that the momentum, as well as the energy, adds up to the correct value of $nLQRmu$ when we sum over all the levels. This is achieved by introducing another selector variable, or Lagrange multiplier, associated with the momentum in the same way that θ is associated with the energy. Its value is afterwards determined in terms of u and turns out to be simply $\exp(u/\boldsymbol{k}T)$ per molecule, and so we must introduce into each term of the partition function a factor $\exp M_x u/\boldsymbol{k}T$, where M_x is the momentum parallel to u for the corresponding level.

Using these results we find for the classical partition function per molecule

$$f = \int_{-\infty}^{+\infty} dp \sum_q \sum_r \exp\left[\frac{-h^2}{8\pi^2 m\boldsymbol{k}T}\left(p^2 + \frac{\pi^2 q^2}{Q^2} + \frac{\pi^2 r^2}{R^2}\right) + \frac{hp}{2\pi}\frac{u}{\boldsymbol{k}T}\right]$$

since $hp/2\pi$ is the momentum in the x-direction according to (15). Performing the integration over p, and replacing the summations over q and r by integrations in the usual way, we find, for an assembly of N molecules

$$N \log f = N\left(\text{constant} + \log(QR) + \frac{3}{2}\log T + \frac{mu^2}{2\boldsymbol{k}T}\right). \qquad \ldots\ldots(15)$$

Applying (4) to an assembly consisting of the molecules within a volume LQR, and allowing Q to vary, (4) becomes

$$P = \frac{N\boldsymbol{k}T}{LR}\frac{\partial}{\partial Q}(\log f) = \frac{\boldsymbol{k}T}{LR}\left(\frac{N}{Q} + \frac{m}{2\boldsymbol{k}T}\frac{d}{dQ}(Nu^2)\right) = n\boldsymbol{k}T - \frac{nmu^2}{2} \qquad \ldots\ldots(16)$$

where N may vary with Q, but n, the number of atoms per unit volume, and Nu, the total rate of transfer of matter, are held constant during the change in Q; (16) is just Bernoulli's expression for the pressure. (We are supposing that we make the measurement by letting the wall deform in such a way that the disturbance of the flow is very small, which is consistent with the above assumptions.) Thus we have verified that the definition (4) of local pressure leads to consistent results in a simple case.

This process can be extended to quantum statistics without difficulty, and a similar treatment is formally possible for the many-body problem if we assume that we know the total energy and total momentum associated with each level of the whole assembly. The introduction of the extra selector variable enables us to decide the relative weights to be given in a steadily streaming assembly to a given level and to the level of equal energy but reversed momentum that are always weighted equally for an assembly in a closed vessel. Our treatment as it stands can be applied, e.g., to a vessel in the form of a sector of a cylinder or sphere, it being only necessary to fix our attention on a region cut off between radii R and $R+dR$, to replace the functions (14) by the corresponding solutions in the cylindrical and spherical cases, and then to calculate the partition function as before under the restraint that we specify the total momentum along the bisector of the angle of the sector. (This is easily seen to lead to expressions equivalent to (15) and (16) in the limit of large radius.) The relationship between momentum and flux now varies with R, but this introduces no difficulty of principle, our general theory showing that any convenient portion of our vessel may be taken as our 'open assembly' without introducing any inconsistency.

ACKNOWLEDGMENTS

I should like to thank Professor G. S. Rushbrooke for helpful discussions, and the Director, Atomic Weapons Research Establishment for permission to publish this paper.

REFERENCES

FOWLER, R. H., 1936, *Statistical Mechanics*, 2nd Edn (Cambridge : University Press); 1938, *Proc. Camb. Phil. Soc.*, **34**, 382.
MASSIGNON, D., 1955, *C. R. Acad. Sci., Paris*, **241**, 1381.
TEMPERLEY, H. N. V., 1952, *Proc. Phys. Soc.* A, **65**, 490.
ZILSEL, P. R., 1953, *Phys. Rev.*, **92**, 1106.

Some Contributions to the Theory of Electrical Conductivity, Thermal Conductivity and Thermoelectric Power in Semiconductors†

By J. E. PARROTT

Research Laboratory, Associated Electrical Industries, Ltd., Aldermaston, Berks.

MS. received 30th January 1957, *and in final form* 1*st March* 1957

Abstract. The theory of transport phenomena in semiconductors is extended to include the effects of 'phonon drag' on electrons and 'electron drag' on phonons. A simple phenomenological theory is developed first, and it is shown that this gives a good semi-quantitative account of the effects under investigation. Commencing with the Boltzmann equation for electrons and phonons, the electrical conductivity, the thermal conductivity and the thermoelectric coefficients are calculated exactly, under certain simplifying assumptions. Where possible, the theoretical results are compared with experimental data and it is shown that general, though not exact, agreement prevails.

§ 1. INTRODUCTION

In the conventional theory of electronic transport phenomena in solids it is assumed that the lattice vibrations which scatter the electrons are in thermal equilibrium. Similarly, in considering heat transport by the lattice vibrations, it is assumed that the electrons which scatter these are in equilibrium. It had been realized for some time that neither of these assumptions was true, but it was considered that the effects of this neglect were likely to be small. Recent experimental work on the thermoelectric power of semiconductors by Frederikse (1953) and Geballe and Hull (1954, 1955) has indicated that this may not always be the case.

Attention was first drawn to these effects by Sommerfeld and Bethe (1933), and an attempt to extend the theory was made by Makinson (1938) in his work on the thermal conductivity of metals. A few years later, Gurevich (1945) applied the same ideas to the thermoelectric effects in metals and Frederikse (1953) employed a simple model to explain his measurements of the thermoelectric power of germanium. Herring (1954 a, b) has given a more detailed analysis of the thermoelectric effects in semiconductors, and ter Haar and Neaves (1955) carried out a similar calculation for metals. A feature of these calculations was that no attempt was made to ensure a symmetrical treatment of the electron and phonon systems. This was shown by Sondheimer (1956) to give results for the thermoelectric coefficients which did not obey the Kelvin relations and which were therefore erroneous. He has further shown that the symmetric treatment must give results in accordance with the Kelvin relations if the model of the kinetic processes obeys the principle of microscopic reversibility.

---

† A brief account of this work was given at the International Colloquium on Semiconductors and Phosphors at Garmisch-Partenkirchen, August 1956.

In this paper, care has been taken to maintain the symmetry of the electron–phonon interactions, at least until the effects of neglecting it have been shown to be small. An orthodox type of Boltzmann equation is used to solve the semiconductor case, in contrast to the method of Herring which uses the concept of momentum transfer. Many of Herring's results are confirmed in this way and, in addition, the effects of phonon and electron disequilibrium on the electrical and thermal conductivities are calculated. These latter effects have been very little considered hitherto. The solution of the problem in exact terms has considerable mathematical complexity, so a phenomenological theory is first developed to give approximate results where no exact solution can be obtained.

In this paper, all the calculations refer to electrons. However, it is always easy to obtain the corresponding results for holes by suitable changes of sign.

§ 2. PHENOMENOLOGICAL THEORY

The equations of phonon–electron interaction, which is described in later sections, are very difficult to solve exactly. For this reason, it is worth while developing first an approximate phenomenological theory which is straightforward to apply. This theory can be used to calculate the thermoelectric power and electrical conductivity, but cannot give the electronic thermal conductivity, since there is no mass movement of electrons in this case.

The basic concept of the phenomenological theory is that if particles of a first type are scattered by particles of a second type, the state into which the first particles tend to be scattered is not the equilibrium state but a state in which they have the average velocity of the particles of the second type. In calculating the average velocity referred to above, the particles of the second type are weighted according to their power of scattering the first type of particle. Another way of expressing this is to say that to an observer moving with the drift velocity of second-type particles, the first-type particles would appear to be scattered into a state at rest in his frame of reference. We use this principle to modify the conventional phenomenological equation of motion of electrons in solids; this is:

$$\frac{dv_e}{dt}=\frac{F}{m}-\frac{v_e}{\tau_e}, \qquad \ldots\ldots(1)$$

where v_e is the average electron velocity, F is the force acting on the electrons, m is their effective mass and τ_e the relaxation time for electrons. The form of the equation shows that the scattering mechanism tends to send the electrons into a state in which they are at rest, so that the 'particles' causing the scattering must be at rest themselves. Furthermore, the use of this equation implies that τ_e is independent of the energy of the electrons; in practice, one has to employ some suitable average value of τ_e. For the stationary state the equation has a solution $v_e=F\tau_e m^{-1}$.

We introduce a moving scattering mechanism, i.e. the phonons, by adding to (1) a term $-(v_e-v_p)\tau_{ep}^{-1}$, where v_p is the average phonon velocity and τ_{ep} the relaxation time for electrons scattered by phonons. The introduction of the additional unknown v_p means that we have to use a similar equation to (1) for the phonons if we are to solve the problem. We also wish to include in F the effects of temperature gradients as well as of electric fields. Inspection of the

Boltzmann equation (discussed in §3) suggests that the suitable expression for F is

$$F=eE+T\frac{\partial}{\partial x}\left(\frac{\zeta_0}{T}\right)+\alpha k\frac{\partial T}{\partial x}, \qquad \ldots\ldots(2)$$

where E is the electric field, ζ_0 the chemical potential of the electrons (the 'Fermi' energy) and α is a numerical constant of order unity depending in particular on the energy dependence of τ_e; for lattice scattering it is equal to two. It is convenient to write (2) somewhat differently:

$$F=\frac{\partial\zeta}{\partial x}-Q_0 e\frac{\partial T}{\partial x}, \qquad \ldots\ldots(3)$$

where Q_0 is the conventional thermoelectric power expression $(\zeta_0-\alpha kT)/eT$, and ζ is the electrochemical potential. The differences in this correspond to the externally measured electromotive force. Using this formulation, the equation of motion of the electrons in the stationary state is

$$\frac{1}{m}\left(\frac{\partial\zeta}{\partial x}-Q_0 e\frac{\partial T}{\partial x}\right)=-\frac{v_e}{\tau_{ei}}-\frac{(v_e-v_p)}{\tau_{ep}}. \qquad \ldots\ldots(4)$$

Here we have introduced τ_{ei}, the relaxation time for scattering by *stationary* impurities.

The phonon Boltzmann equation suggests to us the corresponding phenomenological equation for phonon motion. Corresponding to F/m we have $(s^2/T)\partial T/\partial x$, where s is the velocity of sound, so that the equation is

$$\frac{s^2}{T}\frac{\partial T}{\partial x}=-v_p\left(\frac{1}{\tau_{pp}}-\frac{1}{\tau_{pb}}\right)-\frac{(v_p-v_e)}{\tau_{pe}}. \qquad \ldots\ldots(5)$$

Here, τ_{pp} is the relaxation time for scattering by other phonons, τ_{pb} for scattering by the crystal boundaries, and τ_{pe} for scattering by the electrons moving with velocity v_e.

The thermoelectric coefficients can be calculated from (4) and (5) in two ways, either as the thermoelectric power Q or as the Peltier coefficient Π. These two methods are described by Herring (1954 b) as the Q approach and the Π approach. The Q approach involves the solution of (4) and (5) for $\partial\zeta/\partial x$ under the condition v_e is zero. This is quite straightforward and gives us

$$\frac{\partial\zeta}{\partial x}=e\left(Q_0-\frac{\tau_p}{\tau_{ep}}\frac{ms^2}{eT}\right)\frac{\partial T}{\partial x}, \qquad \ldots\ldots(6)$$

i.e.

$$Q=Q_0-\frac{\tau_p}{\tau_{ep}}\frac{ms^2}{eT}. \qquad \ldots\ldots(7)$$

Here we have written τ_p for the total phonon relaxation time obtained by reciprocal addition of τ_{pp}, τ_{pb} and τ_{pe}. An expression rather like this was derived by the author in an earlier paper (Parrott 1954), and is the same as that derived by different arguments by Frederikse (1953) and Herring (1954 b). The problem in utilizing this expression lies in choice of appropriate expressions for τ_p and τ_{ep}. This question will be returned to later after discussion of the Π approach.

The Peltier coefficient is calculated when $\partial T/\partial x$ is zero and we have a finite electric current nev_e. We then calculate the average phonon velocity and multiply this by the average phonon energy, giving us the heat current carried by the phonons. Division by the electric current then gives the phonon contribution to the Peltier coefficient. The phonon velocity v_p is found to be $f_{pe}v_e$, where f_{pe} is the fraction of the phonon scattering due to electrons ($f_{pe}=\tau_p/\tau_{pe}$). To determine the energy carried, we say that only those longitudinal modes of vibration whose wave vectors are less than the average electron wave vector interact with the electrons and hence have the velocity v_p, and that each of these modes has its classical energy kT. If λ is the average electron wavelength $\hbar(2mkT)^{-1/2}$, then the number of such modes is $(6\pi^2\lambda^3)^{-1}$ and the heat transported is

$$w=\frac{f_{pe}v_e kT}{6\pi^2\lambda^3} \qquad \ldots\ldots(8)$$

whence

$$\Pi=\Pi_0-\frac{f_{pe}kT}{6\pi^2 ne\lambda^3}. \qquad \ldots\ldots(9)$$

For small values of electron density, f_{pe} is proportional to n and the phonon contribution is independent of n. However, as n increases, f_{pe} tends towards unity and we get an inverse dependence on n. This is the saturation effect by which the associated heat current increases less rapidly than the electric current.

The Kelvin relations impose a relation between (7) and (9). This requirement turns out to be an equation relating τ_{pe} and τ_{ep} thus:

$$\frac{\tau_{pe}}{\tau_{ep}}=\frac{\hbar^2}{4\pi^2 nm^2s^2\lambda^5}. \qquad \ldots\ldots(10)$$

The calculations to be described later (equations (28) and (30)) give for τ_{pe} and τ_{ep} expressions such that

$$\frac{\tau_{pe}}{\tau_{ep}}=\frac{\hbar^2}{4\pi^{3/2}nm^2s^2\lambda^5}\frac{x}{y}\exp(y^2), \qquad \ldots\ldots(11)$$

where x and y are reduced wave numbers of electrons and phonons.† If we put $x=y$ and $\exp(-y^2)$ equal to $\pi^{1/2}$, we get complete agreement between the equations (10) and (11). The expression (10) is really derived from the assumption that electrons interact most strongly with phonons of the same wavelength, so that this assumption has played the part of the principle of microscopic reversibility in the phenomenological theory.

In order to get useful expressions from (7) or (9), we need to find some valid expression for τ_p. Herring (1954 a) has shown that for longitudinal phonons of long wavelength at temperatures considerably below the Debye temperature, phonon–phonon scattering gives a relaxation time

$$\tau_{pp}^{-1}=BT^4y^2 \qquad \ldots\ldots(12)$$

where B is a constant. For boundary scattering, we must have approximately

$$\tau_{pb}^{-1}=sL^{-1}, \qquad \ldots\ldots(13)$$

where L is the linear dimension of the crystal.

In the case of a very dilute semiconductor (when boundary scattering can be neglected), $\tau_p^{-1}\simeq BT^4y^2$. Then using the expression for the electron mobility for lattice scattering $\mu_0=e\tau m^{-1}$ we find, putting $y=1$,

$$Q=Q_0-\frac{s^2}{\mu_0 BT^5}. \qquad \ldots\ldots(14)$$

---

† These reduced wave numbers are defined on p. 595.

This closely corresponds with the exact expression which can be calculated in this case.

For the case where boundary scattering predominates, we find

$$Q = Q_0 - \frac{Ls}{\mu_0 T}. \qquad \ldots\ldots(15)$$

This is identical with the exact solution, probably because in this case the relaxation time really is independent of the phonon energy.

Similarly, our analysis suggests that in the case where electrons do nearly all the scattering of the lattice waves so that f_{pe} is unity, we find:

$$Q = Q_0 - \frac{k}{2\pi^3 ne\lambda^3}. \qquad \ldots\ldots(16)$$

Actually, this result is rather dubious, and the case of extreme saturation which it represents will be more fully discussed later. The case of incipient saturation should be accurately given by this simple analysis, however, and we find

$$Q = Q_0 - \frac{s^3}{\mu_0 BT^5}\left(1 - \frac{2\pi^3 nems^2\lambda^5}{BT^4\mu_0 k^3}\right) \qquad \ldots\ldots(17)$$

when the second term in the bracket is small compared with unity.

The phenomenological theory also enables us to calculate the conductivity changes produced by phonon drag. The solution of (4) and (5) in the isothermal case shows that the mobility is given by

$$\mu = \mu_0(1 - f_{ep}f_{pe})^{-1} \qquad \ldots\ldots(18)$$

where f_{ep} is the fraction of electrons scattered by phonons. This shows the interesting fact that as f_{ep} and f_{pe} approach unity, the mobility increases without limit. This is because, if the lattice vibrations are not scattered otherwise than by electrons, there is no mechanism for disposing of the momentum fed into the system by the electric field, and hence the phonons and electrons move more and more rapidly. We shall see below that (18) is in good agreement with the exact solution obtained from the Boltzmann equation.

Unfortunately, the phenomenological equation cannot be used to calculate the electronic thermal conductivity, since it concerns itself only with mass motion of electrons. However, by examining (5) we can see that the effect of electron drag on the lattice conductivity is zero in a first approximation. When $\boldsymbol{v}_e$ is zero, as it is on average in a thermal conductivity experiment, the value of $\boldsymbol{v}_p$ is just what would have been obtained if electron drag had been ignored.

It will thus be seen that a very simple theory can give mathematical expression to the effects produced by electron and phonon disequilibrium as scattering processes. This formulation cannot be regarded as really satisfactory, however, until its relationship to the more correct theory, based on the Boltzmann equation, has been explored.

§ 3. MATHEMATICAL STATEMENT OF THE PROBLEM

The general method to be employed in this section is: firstly, to express the general equations given by Sondheimer in terms applicable to non-degenerate semiconductors; secondly, to extend this statement to include other mechanisms which scatter electrons and phonons; thirdly, to express the departure from equilibrium of the electrons and phonons in terms of a pair of integral equations, and fourthly, to show how these expressions can be used to calculate the electrical conductivity, the thermoelectric coefficients, and the thermal conductivity.

Sondheimer considers together the Boltzmann equations for electrons and phonons when the only scattering of each is by the other. It should be stated here that such a restriction will prevent finite solutions for the current from being obtained, since there is no process provided by which the momentum given to the electrons by the electric field can be dissipated. With this restriction, the rate of change of the electron and phonon distribution functions f and N due to scattering is expressed in terms of four integral operators, $\mathscr{L}_1$, $\mathscr{L}_2$, $\mathscr{L}_3$ and $\mathscr{L}_4$, which contain the transition probabilities from electron–phonon interaction. We have

$$\left[\frac{\partial f}{\partial t}\right]_{\text{coll}} = -\mathscr{L}_1(\Phi) - \mathscr{L}_2(\Psi), \qquad \ldots\ldots(19)$$

$$\left[\frac{\partial N}{\partial t}\right]_{\text{coll}} = -\mathscr{L}_3(\Phi) - \mathscr{L}_4(\Psi), \qquad \ldots\ldots(20)$$

where Φ and Ψ are defined by

$$f = f_0 - \Phi\frac{\partial f_0}{\partial \epsilon} \quad \ldots\ldots(21); \qquad N = N_0 - \frac{\Psi}{kT}\frac{\partial N_0}{\partial z}. \quad \ldots\ldots(22)$$

Here f_0, N_0 are the thermal equilibrium distribution functions, ϵ is the electron energy, z is $\hbar sq/kT$, where s is the velocity of sound in the crystal under consideration, and q is the phonon wave vector. $\mathscr{L}_1$, $\mathscr{L}_2$, $\mathscr{L}_3$, $\mathscr{L}_4$ can be explicitly evaluated for semiconductors using conventional methods such as those described by Wilson (1953). The usual assumptions are made, in particular that: (i) The electron statistics are non-degenerate, so $f_0 = \exp\{-(\epsilon - \zeta_0)/kT\}$, where ζ_0 is the Fermi energy. (ii) Since only the long-wavelength phonons interact with the electrons, we can put $N_0 = z^{-1}$; this assumption will be valid down to all but the very lowest temperatures. (iii) The phonon energies involved in electron–phonon encounters are so small that from the viewpoint of the electrons the collisions are elastic. This will hold down to a few degrees Kelvin.

It is convenient to express the results in terms of some new variables. We define a length $\lambda = \hbar(2mkT)^{-1/2}$, which is the wavelength of a thermal electron, and then proceed to introduce two reduced variables. In place of the electron wave vector $\mathbf{k}$, we shall use $x = \lambda\mathbf{k}$, so that, for example, the energy is given by x^2kT, and in place of phonon wave vector $\mathbf{q}$ we shall use $y = \lambda\mathbf{q}/2$. In terms of λ it is easy to show that $\exp(\zeta_0/kT) = 4\pi^{3/2}\lambda^3 n$, where n is the density of electrons. We introduce $c(x)$ defined by $\Phi = k_x c(x)$, and $b(y)$ defined by $\Psi = q_x b(y)$. Lastly, we express the electron–lattice coupling in terms of the electron mobility for pure lattice scattering, μ_0. Using these quantities, we find

$$\mathscr{L}_1(\Phi) = \frac{32\pi k_x c(x) ne\lambda^5 x \exp(-x^2)}{3\mu_0 \hbar^2}, \qquad \ldots\ldots(23)$$

$$\mathscr{L}_2(\Psi) = \frac{128\pi k_x ne\lambda^5 \exp(-x^2)}{3\mu_0 \hbar^2 x^3}\int_0^x y^3 b(y)\,dy, \qquad \ldots\ldots(24)$$

$$\mathscr{L}_3(\Phi) = \frac{4\pi q_x ne\lambda^5}{3\mu_0 \hbar^2 y}\int_y^\infty \exp(-x^2)x\,c(x)\,dx, \qquad \ldots\ldots(25)$$

$$\mathscr{L}_4(\Psi) = \frac{2\pi q_x b(y) ne\lambda^5 \exp(-y^2)}{3\mu_0 \hbar^2 y}. \qquad \ldots\ldots(26)$$

Here, the middle pair of equations (24) and (25) represent the interactions usually neglected. The limits in (24) arise because no lattice wave with a wave number greater than $2k$ can interact with an electron of wave number k, and conversely with equation (25).

Before proceeding further, we have to expand equations (19) and (20) to include other scattering mechanisms. If τ_e is the electron relaxation time, in general dependent on x, then the general expression is

$$\left[\frac{\partial f}{\partial t}\right]_{\mathrm{coll}} = \frac{8\pi^{3/2}k_x c(x) nm\lambda^5 \exp(-x^2)}{\hbar^2 \tau_e}. \qquad \ldots\ldots(27)$$

For electrons we have (i) the relaxation time for electron–phonon collisions τ_{ep} which is, by definition,

$$\tau_{\mathrm{ep}}{}^{-1} = \frac{4ex}{3\pi^{1/2}\mu_0 m} \qquad \ldots\ldots(28)$$

and (ii) the relaxation time for impurity scattering τ_{ei} which will differ according to whether the impurity is ionized or neutral. In the former case, $1/\tau_{\mathrm{ei}} \propto x^{-3}$, and in the latter it is independent of x. These cases have been investigated by Conwell and Weisskopf (1950), Brooks (1951) and by Erginsoy (1950).

In the phonon case, if τ_{p} is a general phonon relaxation time, then

$$\left[\frac{\partial N}{\partial t}\right]_{\mathrm{coll}} = -\frac{q_x b(y)}{8ms^2\tau_{\mathrm{p}}y^2}. \qquad \ldots\ldots(29)$$

By comparing this with equation (26) we see that the phonon relaxation time for electron scattering is given by

$$\tau_{\mathrm{pe}}{}^{-1} = \frac{16\pi ne\lambda^5 ms^2 y \exp(-y^2)}{3\mu_0\hbar^2}. \qquad \ldots\ldots(30)$$

This relaxation time has the expected properties of being inversely proportional to the number density of electrons and of increasing exponentially with the phonon wave number because of the relatively few electrons of high energy.

The total relaxation times can be obtained by reciprocal additions of the component relaxation times. Thus, we can write down the total values of $[\partial f/\partial t]_{\mathrm{coll}}$ and $[\partial N/\partial t]_{\mathrm{coll}}$ as follows:

$$\left[\frac{\partial f}{\partial t}\right]_{\mathrm{coll}} = -\frac{8\pi^{3/2}k_x nm\lambda^5 \exp(-x^2)}{\hbar^2}\left[c(x)\left(\frac{1}{\tau_{\mathrm{ep}}}+\frac{1}{\tau_{\mathrm{ei}}}\right) - \frac{4}{\tau_{\mathrm{ep}}x^4}\int y^3 b(y)\,dy\right], \qquad \ldots\ldots(31)$$

$$\left[\frac{\partial N}{\partial t}\right]_{\mathrm{coll}} = -\frac{q_x}{8ms^2y^2}\left[b(y)\left(\frac{1}{\tau_{\mathrm{pe}}}+\frac{1}{\tau_{\mathrm{pp}}}+\frac{1}{\tau_{\mathrm{pb}}}\right) - \frac{2\exp(y^2)}{\tau_{\mathrm{pe}}}\int \exp(-x^2)x\,c(x)\,dx\right]. \qquad \ldots\ldots(32)$$

The integrals describe respectively the phonon drag on the electrons and the electron drag on the phonons.

We have next to write down the variation of f and N due to electric fields and temperature gradients. From, for example, Wilson (1953) we write down:†

$$\left[\frac{\partial f}{\partial t}\right]_{\mathrm{drift}} = -\frac{\hbar k_x}{m}\frac{\partial f_0}{\partial \epsilon}\left\{eE + T\frac{\partial}{\partial x}\left(\frac{\zeta_0}{T}\right) + \frac{\epsilon}{T}\frac{\partial T}{\partial x}\right\}. \qquad \ldots\ldots(33)$$

---

† Here we have for convenience also used x as a spatial variable. This should not cause confusion, as it appears only in the form $\partial/\partial x$ and as the subscript in k_x, q_x.

In terms of the electrochemical potential ζ and using the various reduced variables described earlier, we express (33) as:

$$\left[\frac{\partial f}{\partial t}\right]_{\text{drift}} = -\frac{8\pi^{3/2}k_x\lambda^3 n\exp(-x^2)}{\hbar}\left(\frac{\partial\zeta}{\partial x} - \frac{(\zeta_0 - x^2kT)}{T}\frac{\partial T}{\partial x}\right). \quad \ldots\ldots(34)$$

Again, from Wilson, we find for the rate of change of N due to a temperature gradient:

$$\left[\frac{\partial N}{\partial t}\right]_{\text{drift}} = \frac{sq_x}{q}\frac{\partial N_0}{\partial T}\frac{\partial T}{\partial x}. \quad \ldots\ldots(35)$$

In terms of the reduced variables, this becomes

$$\left[\frac{\partial N}{\partial t}\right]_{\text{drift}} = \frac{q_x k\lambda^2}{4\hbar^2 y^2}\frac{\partial T}{\partial x}. \quad \ldots\ldots(36)$$

We are now in a position to write down the Boltzmann equations for both electrons and phonons. The resulting pair of simultaneous integral equations for $c(x)$ and $b(y)$ are:

$$c(x) = -\frac{\hbar\tau_e}{m}\left(\frac{\partial\zeta}{\partial x} - \frac{(\zeta_0 - x^2kT)}{T}\frac{\partial T}{\partial x}\right) + \frac{4f_{\text{ep}}}{x^4}\int_0^x y^3 b(y)\,dy \quad \ldots\ldots(37)$$

and

$$b(y) = -\frac{\hbar s^2\tau_p}{T}\frac{\partial T}{\partial x} + 2f_{\text{pe}}\exp(y^2)\int_y^\infty \exp(-x^2)x\,c(x)\,dx. \quad \ldots\ldots(38)$$

By substitution of (38) in (37) and (37) into (38), we obtain two double integral equations for $c(x)$ and $b(y)$ separately. These are:

$$c(x) - \frac{8f_{\text{ep}}}{x^4}\int_0^x f_{\text{pe}}\exp(y^2)y^3\int_y^\infty \exp(-x^2)x\,c(x)\,dx\,dy$$
$$= -\frac{\hbar\tau_e}{m}\left\{\frac{\partial\zeta}{\partial x} - \frac{1}{T}\frac{\partial T}{\partial x}\left[\zeta_0 - x^2kT - \frac{4f_{\text{ep}}ms^2}{\tau_e x^4}\int_0^x \tau_p y^3\,dy\right]\right\} \quad \ldots\ldots(39)$$

and

$$b(y) - 8f_{\text{pe}}\exp(y^2)\int_y^\infty f_{\text{ep}}\exp(-x^2)x^{-3}\int_0^x y^3 b(y)\,dy\,dx$$
$$= -\frac{\hbar s^2\tau_p}{T}\frac{\partial T}{\partial x} - \frac{2\hbar f_{\text{pe}}\exp(y^2)}{m}\left\{\left[\frac{\partial\zeta}{\partial x} - \frac{\zeta_0}{T}\frac{\partial T}{\partial x}\right]\int_y^\infty \exp(-x^2)x\tau_e\,dx\right.$$
$$\left. + k\frac{\partial T}{\partial x}\int_y^\infty \exp(-x^2)x^3\tau_e\,dx\right\}. \quad \ldots\ldots(40)$$

Equations (39) and (40) are the general mathematical statement of the problem we are considering, and it will be seen that their exact solution is a problem of considerable complexity except in simplified cases. Their general significance is, however, fairly clear. In the case of both $c(x)$ and $b(y)$, the second term on the left-hand side represents a second-order effect of the type described by saying that it is produced on the electrons by the changes in the phonon distribution caused by the disequilibrium of the electrons. It is the term responsible for the effects of phonon drag on the electric conductivity. On the right-hand side of the equations, we have first of all the normal drift terms and then, under the integral signs, the terms representing the first-order phonon drag on the electrons and electron drag on phonons respectively. These will normally be the terms of greatest importance in calculating the thermoelectric effects. Finally, it may be observed that, since either $c(x)$ or $b(y)$ can be expressed in terms of the other, the compound system phonons–electrons can be regarded as a true unity describable by a single function.

The next stage in the process is to calculate the electrical and thermal currents from the distribution perturbations $c(x)$ and $b(y)$; the general mathematical expression of this is described by Wilson (1953). To calculate the electrical conductivity or, what is more significant in the case of semiconductors, the electron mobility, we calculate the current in an isothermal conductor. Equation (39) simplifies considerably in this case and the $c(x)$ derived is used in

$$\mu = \frac{8}{3\pi^{1/2}\hbar E}\int_0^\infty \exp(-x^2)x^4 c(x)\,dx. \qquad \ldots\ldots(41)$$

The thermoelectric power Q is determined by the value of

$$\left(\frac{1}{e}\frac{\partial \zeta}{\partial x}\right)\Big/\frac{\partial T}{\partial x}$$

when the electric current is zero. Since the current is proportional to $\int \exp(-x^2)x^4 c(x)\,dx$, the condition for zero current can be written formally as

$$\int\Bigg\{\exp(-x^2)x^4 \times\left[\tau_e\left\{\frac{\partial \zeta}{\partial x}-\frac{1}{T}\frac{\partial T}{\partial x}(\zeta_0 - x^2kT)-\frac{4f_{\mathrm{ep}}ms^2}{\tau_e x^4}\int \tau_{\mathrm{p}} y^3\,dy\right\}\right](1-\mathscr{S})^{-1}\Bigg\}\,dx = 0, \qquad \ldots\ldots(42)$$

where $\mathscr{S}$ is the integral operator acting on $c(x)$ on the left-hand side of (39). The solution of this equation gives Q. The Peltier coefficient, on the other hand, is calculated from the heat transport in an isothermal conductor. The portion of this attributable to electrons is given by

$$W_e = -\frac{8n}{3\pi^{1/2}\hbar}\left[\zeta_0\int_0^\infty \exp(-x^2)x^4 c(x)\,dx - kT\int_0^\infty \exp(-x^2)x^6 c(x)\,dx\right] \qquad \ldots\ldots(43)$$

and that attributable to phonons by

$$w_{\mathrm{p}} = \frac{64ms^2}{3\pi^2\hbar\lambda^2}\int_0^{y_1} y^4 b(y) N_0(N_0+1)\,dy, \qquad \ldots\ldots(44)$$

where $y_1 = \frac{1}{2}(kT/ms^2)^{1/2}\theta/T$. If we divide these expressions by the expression for the current, we get the Peltier coefficient consisting of three terms corresponding to transport of electron potential energy, electron kinetic energy and phonon energy.

If, on the other hand, we take the value of $c(x)$ calculated in the zero-current condition, use this to determine $b(y)$ and then insert them into (43) and (44), we shall obtain the heat transport in a temperature gradient. Division by $\partial T/\partial x$ will give the thermal conductivity.

The calculation of the electrical and thermal conductivity and of the thermoelectric coefficients will be carried out for certain fairly simple cases in the following sections.

§ 4. THE ELECTRICAL CONDUCTIVITY

It is clear that the true electrical conductivity must differ from that calculated from the assumption that the lattice vibrations are in equilibrium. The movement of the electrons is transmitted in part to the phonons, with a consequent reduction in their power of scattering the electrons. The effect is limited by two factors: the scattering of phonons by other mechanisms, which reduces their departure from equilibrium, and the scattering of electrons by other mechanisms, which will become relatively more important if there is a reduction in lattice scattering. Without these factors, the resistivity of a conductor would go to zero.

To treat the problem mathematically, we first solve (39) for $c(x)$ in isothermal conditions, and then use this $c(x)$ to solve (41). The appropriate form of (39) is

$$c(x)-\frac{8f_{\mathrm{ep}}}{x^4}\int_0^x f_{\mathrm{pe}}y^3\exp(y^2)\int_y^\infty \exp(-x^2)x\,c(x)\,dx\,dy=\frac{e\hbar E f_{\mathrm{ep}}\tau_{\mathrm{ep}}}{m}. \qquad \ldots\ldots(45)$$

Using the $\mathscr{S}$-notation introduced in the last section, we can write down a formal solution of (45) as

$$c(x)=\frac{e\hbar E}{m}\frac{f_{\mathrm{ep}}\tau_{\mathrm{ep}}}{1-\mathscr{S}}=\frac{e\hbar E}{m}[f_{\mathrm{ep}}\tau_{\mathrm{ep}}+\mathscr{S}f_{\mathrm{ep}}\tau_{\mathrm{ep}}+\mathscr{S}^2f_{\mathrm{ep}}\tau_{\mathrm{ep}}+\ldots], \qquad \ldots\ldots(46)$$

if $\mathscr{S}$ is less than unity. Equation (46) represents an iterative process for obtaining $c(x)$ as a series of terms. Similarly, we can express the mobility itself as a sum of terms:

$$\mu=\mu'+\mu''+\mu'''+\ldots, \qquad \ldots\ldots(47)$$

where

$$\mu'=2\mu_0\int_0^\infty \exp(-x^2)x^3f_{\mathrm{ep}}\,dx,$$

$$\mu''=16\mu_0\int_0^\infty \exp(-x^2)f_{\mathrm{ep}}\int_0^x \exp(y^2)y^3f_{\mathrm{pe}}\int_y^\infty \exp(-x^2)f_{\mathrm{ep}}\,dx\,dy\,dx,$$

$$\mu'''=128\mu_0\int_0^\infty \exp(-x^2)f_{\mathrm{ep}}\int_0^x \exp(y^2)y^3f_{\mathrm{pe}}\int_y^\infty \exp(-x^2)x^{-3}f_{\mathrm{ep}}\int_0^x \exp(y^2)$$
$$\times y^3f_{\mathrm{pe}}\int \exp(-x^2)f_{\mathrm{ep}}\,dx\ldots dx.$$

It will be seen that if f_{ep} is unity, then the first term $\mu'=\mu_0$, the normal lattice scattering mobility. It will be convenient to introduce a notation $\langle f_{\mathrm{ep}}\rangle$ for the average of f_{ep} such that

$$\langle f_{\mathrm{ep}}\rangle=2\int_0^\infty \exp(-x^2)x^3f_{\mathrm{ep}}\,dx. \qquad \ldots\ldots(48)$$

Then we can simply write $\mu'=\langle f_{\mathrm{ep}}\rangle\mu_0$. The second term will depend on both f_{ep} and f_{pe}, and if either of these is small, μ'' is also small and μ''' can be ignored completely. For the case where f_{ep} and f_{pe} are both unity, we should find that the series of terms (47) does not converge; μ'' turns out to be

$$\mu''=2\left(\pi-\frac{8}{3}\right)\mu_0\simeq 0{\cdot}95\mu_0 \qquad \ldots\ldots(49)$$

in this case. The calculation of μ''' was not proceeded with, but the value of μ'' is not inconsistent with the expected infinite value of μ in this case.

A more practical case is that when f_{ep} is unity but $f_{\mathrm{pe}}\simeq\tau_{\mathrm{pp}}/\tau_{\mathrm{pe}}$, i.e. when the scattering of phonons by other phonons predominates. The value of $\tau_{\mathrm{pp}}/\tau_{\mathrm{pe}}$ is given, using (28) and (30), by

$$\frac{\tau_{\mathrm{pp}}}{\tau_{\mathrm{pe}}}=\frac{16\pi ne\lambda^5ms^2}{3\mu\hbar^2BT^4}\frac{\exp(-y^2)}{y}=\frac{\exp(-y^2)}{\alpha y}, \qquad \ldots\ldots(50)$$

which also defines α. Then

$$\mu'' = \frac{8\pi^{1/2}\mu_0}{\alpha}\int_0^\infty \exp(-x^2)\int_0^x \operatorname{erf}(y)\,y^2\,dy\,dx$$

$$= \frac{8\pi^{3/2}}{3\alpha}\left(1-\frac{5\sqrt{2}}{8}\right)\mu_0 = \frac{0{\cdot}95\mu_0}{\alpha}. \qquad \ldots\ldots(51)$$

If α is large, μ'' need not concern us as it is quadratic in α^{-1}. Then

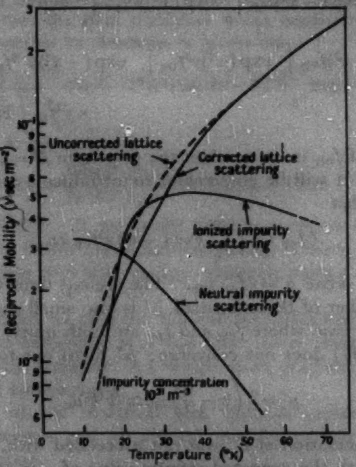
$$\mu = (1+0{\cdot}95/\alpha)\mu_0. \qquad \ldots\ldots(52)$$

Comparing this with equation (18) we see that the agreement is quite good between the two equations if f_{pe} is small and we can expand (18). The effect of boundary scattering of the lattice would be further to diminish μ'' by driving the phonons nearer to their equilibrium distribution.

It is of interest to attempt to calculate the quantitative influence of these effects on the mobility of a semiconductor. This we will do for the case of germanium. It is convenient and not productive of too great an error to put f_{ep} equal to unity in the above equations and regard μ as μ_{L}, the lattice scattering mobility. The total mobility is then obtained by reciprocal addition to μ_{L} of the mobilities derived from ionized and neutral impurity scattering, i.e.

$$\mu^{-1} = \mu_{\mathrm{L}}^{-1} + \mu_{\mathrm{iI}}^{-1} + \mu_{\mathrm{nI}}^{-1}. \qquad \ldots\ldots(53)$$

In the calculation, we use a value of B derived from the analysis of thermoelectric data to be described later. The results are shown in the figure, where we have



Electron scattering processes.

plotted the reciprocal mobilities at low temperatures for germanium, having a donor concentration of 10^{21} m^{-3}. The uncorrected lattice mobility is shown by a broken line and the corrected mobility by a full line. We see that correction disappears at higher temperatures due to the increase of α with T (see (50)), and diminishes at lower temperatures due to the condensation of the conduction electrons on to the impurity centre. This last effect decides the relative importance of ionized and neutral impurity scattering which is also shown. The data used here were drawn from Debye and Conwell (1954) and, following their analysis, an effective mass ratio of 1/8 was used in calculating ionized impurity scattering, and of 1/3 for neutral impurity scattering. It will be seen that the influence of phonon drag on the total scattering is unlikely to exceed 10% and consequently, in view of the other sources of confusion at such low temperatures, e.g. impurity band conduction, it will be difficult to identify in practice.

§ 5. THE THERMOELECTRIC EFFECTS

As stated in § 3, the thermoelectric effects will be calculated in two ways, through the Peltier effect and through the Seebeck effect. A third method, using the Thomson effect, is also possible, but will not be used here. The author has given elsewhere (Parrott 1956) an extensive account of the thermodynamic theory of the thermoelectric effects, analysing the relationship between the electrochemical potential and the Seebeck effect and the connection between the Peltier effect, Thomson effect, and the heat and energy currents in a conductor. This leads to the Kelvin relations between the thermoelectric coefficients and Sondheimer (1956) has further shown that a kinetic theory of the type described must give coefficients obeying the Kelvin relations.

In view of the increasing complexity of the equations an abbreviated notation will be introduced here. We shall write

$$\int_0^\infty \exp(-x^2) g(x) \int_0^x f(y)\, dy\, dx \qquad \text{as} \qquad \langle g(x)[f(y)]\rangle$$

and

$$\int \exp(-x^2) g(x)\, dx \qquad \text{as} \qquad \langle g(x)\rangle.$$

The simplest calculation is undoubtedly that using the Peltier approach, since it is carried out on an isothermal conductor. The heat flow has been written down in (43) and (44) so that the Peltier coefficient is given by the following equations, where it is broken up into electron and phonon contributions Π_e and Π_p:

$$\Pi_e = \frac{\zeta_0}{e} - \frac{kT}{e}\frac{\langle x^6 c(x)\rangle}{\langle x^4 c(x)\rangle}, \qquad \ldots\ldots(54)$$

$$\Pi_p = -\frac{kT}{e}\frac{2}{\pi^{3/2} n\lambda^3}\frac{\langle x\, c(x)[f_{p0} \exp(y^2) y^2]\rangle}{\langle x^4 c(x)\rangle}. \qquad \ldots\ldots(55)$$

The derivation of (55) calls for some explanation. Firstly, equation (44) is modified: (i) by putting N_0 equal to its long-wavelength approximation; this is reasonable since long-wavelength vibrations are the only ones of importance; (ii) the upper limit is made infinite; this has the same justification as (i); (iii) we use (38) to substitute $c(x)$ for $b(y)$, $\partial T/\partial x$ being zero; (iv) we reverse the order of integration as shown by Herring (1954 b), thereby making the calculation simpler.

The role of $c(x)$ in these equations (54), (55) is seen to be not very important, since it occurs in both numerator and denominator throughout. The form of $c(x)$ is modified by the operator $(1-\mathscr{S})^{-1}$, but this can only affect the result through its effect on the x-dependence; mere changes of magnitude would have no effect. For this reason, it will not produce very large errors if $\mathscr{S}$ is ignored altogether. In this approximation, the second term in (54) is unaltered.

Qualitatively, (55) may be expected to show the following features: (*a*) when f_{pe} is very small, it is equal to τ_{pp}/τ_{pe} and therefore proportional to n (this will make Π_p independent of n); (*b*) as f_{pe} grows larger, it is no longer proportional to n and tends to a limit at unity. This results in a steady reduction of Π_p tending towards an inverse proportionality to n. This is the 'saturation' effect (Herring 1954 b), so called because the increase in the magnitude of n is not accompanied by a proportional increase in the heat current dragged along.

Before discussing (54) and (55) further, we will derive the thermoelectric power from (43). Again, the point can be made immediately that the operator $\mathscr{S}$ has only a secondary effect, since it operates equally on $\partial\zeta/\partial x$ and on the other terms. The only effect it produces is through the varying energy dependence of these terms. If we neglect $\mathscr{S}$ then the equation can be written

$$\left(\frac{\partial\zeta}{\partial x}-\frac{\zeta_0}{T}\frac{\partial T}{\partial x}\right)\langle x^3 f_{ep}\rangle+\frac{k\partial T}{\partial x}\langle x^5 f_{ep}\rangle+\frac{16s^2}{3\pi^{1/2}\mu_0 T}\frac{\partial T}{\partial x}\langle f_{ep}[f_{pe}\tau_{pe}y^2]\rangle=0. \qquad \ldots\ldots(56)$$

From this

$$\begin{aligned}Q&=\frac{1}{e}\frac{\partial\zeta}{\partial x}\Big/\frac{\partial T}{\partial x}\\&=\frac{\zeta_0}{eT}-\frac{k}{e}\frac{\langle x^5 f_{ep}\rangle}{\langle x^3 f_{ep}\rangle}-\frac{16s^2}{3\pi^{1/2}\mu_0 T}\frac{\langle f_{ep}(f_{pe}\tau_{pe}y^2)\rangle}{\langle f_{ep}x^3\rangle}.\end{aligned} \qquad \ldots\ldots(57)$$

Remembering that if $\mathscr{S}$ is ignored $c(x)\propto\tau_e\propto f_{ep}x^{-1}$ and that

$$\tau_{pe}=3\mu\hbar^2/16\pi ne\lambda^5 ms^2 y\exp(-y^2),$$

we shall see that equations (54), (55) and (57) give results in accord with the Kelvin relations. However, it should be recalled that (55) was derived from (44) by making certain extra assumptions arising from the preponderant importance of the long-wavelength vibrations. No such extra assumptions are needed in deducing (57) which shows that in fact the assumptions made in obtaining (55) were not only reasonable but necessary for the calculations to be consistent with the basic assumptions discussed in § 3, particularly assumption (ii). This is important when one considers the modifications necessary in dealing with high carrier concentrations which may approach degeneracy. It will be seen to be necessary to lift both assumptions (i) and (ii) in this case, with consequent increased complication of the equations, especially in the semi-degenerate range.

We will now proceed to give more detailed consideration to Π_p, which we most conveniently write as

$$\Pi_p=\frac{kT}{e}\frac{2}{\pi^{3/2}n\lambda^3}\frac{\langle f_{ep}[f_{pe}\exp(y^2)y^2]\rangle}{\langle f_{ep}x^3\rangle}, \qquad \ldots\ldots(58)$$

where

$$f_{pe}=[1+\alpha y\exp(y^2)+\beta y^{-1}\exp(y^2)]^{-1}$$

and

$$\alpha=\frac{3\mu_0\hbar^2BT^4}{16\pi ne\lambda^5 ms^2},\qquad\beta=\frac{3\mu_0\hbar^2}{16\pi ne\lambda^5 msL},\qquad f_{ep}=\left[1+\frac{\tau_{ep}}{\tau_{ei}}\right]^{-1}.$$

It will be readily seen that no analytical solution of the problem in general can be obtained. Further, the integral with $f_{pe}=1$ diverges, showing the breakdown of the basic assumptions. However, the equation can easily be solved for the extreme cases of α very large and β very large. In the first case we find

$$\Pi_p = -\frac{4s^2}{3BT^4\mu_0}, \qquad \ldots\ldots(59)$$

as obtained by Herring (1954 b). This applies to a dilute semiconductor where n is small, with a crystal dimension sufficient to make β very much smaller than α. We shall show that the results on the purest germanium and silicon at not too low temperatures can be explained by the use of (58). Another case is when boundary scattering has the greatest importance and $\beta \gg \alpha$ and $\beta \gg 1$. In this case

$$\Pi_p = -Ls/\mu_0. \qquad \ldots\ldots(60)$$

It should be stated here that this result is only approximate and the reader is referred to Herring's (1954 b) papers for a more exact analysis. Comparison of the phenomenological equations (14) shows that this method gives results in good accord with the exact calculations.

A most important use of (59) is in the calculation of B from experimental data. These results are set out in the table below. The thermoelectric data come from Geballe and Hull (1954, 1955), the mobility data from Morin (1954) and Morin and Maita (1954), and the sound velocity calculated from the elastic constants using the results of Bond *et al.* (1951) and McSkimin *et al.* (1954).

Material	Π_p(v)	μ (m^{2}v^{-1} sec^{-1})	s (m sec^{-1})	B (sec^{-1} deg^{-4})
n-germanium	$-8{\cdot}6\times10^{-2}$	3·5	$5{\cdot}4\times10^{3}$	3·19
p-germanium	$8{\cdot}0\times10^{-2}$	3·8		3·14
n-silicon	$-6{\cdot}2\times10^{-1}$	4·6	$9{\cdot}2\times10^{3}$	0·96
p-silicon	$8{\cdot}0\times10^{-1}$	1·07		3·18

It will be seen that there is a remarkable agreement in the B values except in the case of n-silicon. Agreement is to be expected between n and p samples of the same material since B should not depend on the conductivity type being a measure of lattice anharmonicity. Likewise, since the structures of germanium and silicon are so similar, we might expect that the B values would be similar. The reason for the n-silicon discrepancy may lie in the mobility value used which was extrapolated from high temperatures using a $T^{-2\cdot6}$ formula. If this is due to inter-valley scattering, it might be expected that the temperature variation would be less steep at lower temperatures making the mobility smaller and bringing the B value up somewhat. However, the threefold change needed seems rather large to obtain in this way.

The temperature dependence of Π_p in the purest specimens was found by Geballe and Hull to be $T^{-1\cdot4}$ for n-germanium, $T^{-2\cdot2}$ for p-germanium, $T^{-1\cdot3}$ for n-silicon and $T^{-1\cdot3}$ for p-silicon. According to the theory, we should have $\Pi_p \propto T^{-4-s}$ where s is the power of temperature variation of mobility. Morin (1954) gives values of s as follows: n-germanium $-1{\cdot}6$, p-germanium $-2{\cdot}33$, n-silicon $-2{\cdot}6$ and p-silicon $-2{\cdot}3$. It will be seen that the agreement with the theory is best for n-silicon where, however, as we have seen, the mobility index may in fact not be very accurately known. Both the onset of saturation and boundary scattering will tend to reduce the index, but this will not do for the case of p-germanium where the discrepancy is of different sign.

The value of α is given by

$$\alpha=\frac{B\mu_0 T^{6\cdot5}}{n_s^{\,2}}\left(\frac{m}{m_0}\right)^{3/2}\times 1\cdot10\times10^{18},$$

and that of β by

$$\beta=\frac{\mu T^{2\cdot5}}{Lns}\left(\frac{m}{m_0}\right)^{3/2}\times 1\cdot10\times10^{18}.$$

The value of α was that used in calculating the mobility changes. To give an idea of the numerical orders of magnitude, α is equal to unity in n-germanium at 80°K when the carrier concentration is $2\cdot3\times10^{22}\,\mathrm{m}^{-3}$, in p-germanium when it is $2\cdot2\times10^{22}\,\mathrm{m}^{-3}$, in n-silicon when it is $1\cdot56\times10^{23}\,\mathrm{m}^{-3}$ and in p-silicon $4\cdot6\times10^{22}\,\mathrm{m}^{-3}$. If the sample dimension L is 1 mm, β will be about 20 times smaller at the same temperature and carrier concentration. Due to the more rapid decrease of α with decreasing temperature, α and β will be approximately equal at 30°K; the value of n for which they are both equal to unity will be more than two orders of magnitude less at a few times $10^{20}\,\mathrm{m}^{-3}$. This corresponds to a highly purified specimen which, in the case of germanium, would be almost intrinsic at room temperature.

§ 6. THE THERMAL CONDUCTIVITY

The thermal conductivity, made up as it is of contributions from the electrons and from the lattice waves, will obviously be affected by the phonon–electron drag, but one would expect the effect to be small since, in the condition of zero electron current under which it is measured, the electrons are forced to be on the average at rest. In dilute semiconductors the lattice conductivity preponderates, so that changes in the electron distribution produced by the phonon flow will not greatly affect the thermal conductivity. In semiconductors having a higher electron concentration, some effects might be observable, however. For this reason the problem is worth considering.

The departure from equilibrium of the electrons measured by $c(x)$ can be determined by first solving (42), as was done in the thermoelectric power calculation, and then inserting the value of $\partial\zeta/\partial x$ so obtained in (39). Ignoring the small effect produced by the integral operator $\mathscr{S}$ we find

$$c(x)=\frac{\hbar\tau_e k}{m}\frac{\partial T}{\partial x}\left[\left(\frac{\langle x^6\tau_e\rangle}{\langle x^4\tau_e\rangle}-x^2\right)+\frac{16es^2}{3\pi^{1/2}\mu_0 kT}\left(\frac{\langle x\tau_e(\tau_p y^3)\rangle}{\langle x^4\tau_e\rangle}-\frac{(\tau_p y^3)}{x^3}\right)\right].\quad\ldots(61)$$

The first bracketed term in (61) is the one obtained when phonon drag is neglected. It tells us that the more energetic electrons move down the temperature gradient and the less energetic electrons move up it, giving a normal positive thermal conductivity. The second term is due to phonon drag and in some circumstances gives rise to a reverse drift of the electrons. This occurs if $\tau_p y^3$ is less than proportional to x^3 as will be the case when phonon–phonon scattering predominates. Since the second term will be larger than the first at low temperatures, we may find a $c(x)$ which corresponds to a negative electronic thermal conductivity. The physical reason for this is that the slower electrons are more closely coupled to the phonons than the faster ones and thus tend to be dragged down the temperature gradient. We shall see that there is a similar effect of the electrons on the phonons. The calculated electronic thermal conductivity is

$$\kappa_e = \sigma T\left(\frac{k}{e}\right)^2\left[\frac{\langle x^8\tau_e\rangle}{\langle x^4\tau_e\rangle} - \frac{\langle x^6\tau_e\rangle^2}{\langle x^4\tau_e\rangle^2} + \frac{16es^2}{3\pi^{1/2}\mu_0 kT}\right.$$
$$\left.\times\left(\frac{\langle x^3\tau_e(\tau_p y^3)\rangle}{\langle x^4\tau_e\rangle} - \frac{\langle x^6\tau_e\rangle\langle x\tau_e(\tau_p y^3)\rangle}{\langle x^4\tau_e\rangle^2}\right)\right]. \qquad \ldots\ldots(62)$$

The lattice thermal conductivity is calculated by obtaining $b(y)$ from $c(x)$ using (38) and (61), and inserting $b(y)$ into (44). It conveniently breaks up into two parts

$$\kappa_p = \frac{32s^2k}{3\pi^2}\int y^2\tau_p\,dy, \qquad \ldots\ldots(63)$$

and

$$\kappa_p{}' = \sigma T\left(\frac{k}{e}\right)^2\left[\left(\frac{16es^2}{3\pi^{1/2}\mu_0 kT}\right)\left(\frac{\langle x^3\tau_e(\tau_p y^3)\rangle}{\langle x^4\tau_e\rangle} - \frac{\langle x^6\tau_e\rangle\langle x\tau_e(\tau_p y^3)\rangle}{\langle x^4\tau_e\rangle^2}\right)\right.$$
$$\left.+\left(\frac{16es^2}{3\pi^{1/2}\mu_0 kT}\right)^2\left(\frac{\langle x^{-2}\tau_e(\tau_p y^3)^2\rangle}{\langle x^4\tau_e\rangle} - \frac{\langle x\tau_e(\tau_p y^3)\rangle^2}{\langle x^4\tau_e\rangle^2}\right)\right]. \qquad \ldots\ldots(64)$$

The term κ_p is the normal phonon thermal conductivity and its calculation lies outside our province in this paper. For this reason, we shall not consider it further but merely regard it as a quantity to be added to κ_e and $\kappa_p{}'$. $\kappa_p{}'$ is the electron-drag term divided into two parts: the first is the effect produced on the phonons by the normal electron distribution as far as thermal conductivity is concerned, whilst the second is the effect produced by the phonon-perturbed electron distribution. The first of these terms is, in fact, identical with the second term in (62) and is of the order of eQ_p/k, whilst the second is of order $(eQ_p/k)^2$. Finally, the first term is negative where phonon–phonon scattering predominates, and the second is positive. Thus, we seem to have equal negative contributions to κ_e and $\kappa_p{}'$, suggesting a close connection between κ_e and $\kappa_p{}'$. In fact, $\kappa_p{}'$ can be regarded as an extra electron term, since it depends entirely on electronic motion to produce it. Therefore, we will add equations (62) and (63) together and express it as an effective Wiedemann–Franz–Lorenz (WFL) number:

$$\frac{\kappa_e+\kappa_p{}'}{\sigma T} = \left(\frac{k}{e}\right)^2\left[\left(\frac{\langle x^8\tau_e\rangle}{\langle x^4\tau_e\rangle} - \frac{\langle x^6\tau_e\rangle^2}{\langle x^4\tau_e\rangle^2}\right) - \frac{32es^2}{3\pi^{1/2}\mu_0 kT}\left(\frac{\langle x^6\tau_e\rangle\langle x\tau_e(\tau_p y^3)\rangle}{\langle x^4\tau_e\rangle^2}\right.\right.$$
$$\left.\left.-\frac{\langle x^3\tau_e(\tau_p y^3)\rangle}{\langle x^4\tau_e\rangle}\right) + \left(\frac{16es^2}{3\pi^{1/2}\mu_0 kT}\right)^2\left(\frac{\langle x^{-2}\tau_e(\tau_p y^3)^2\rangle}{\langle x^4\tau_e\rangle} - \frac{\langle x\tau_e(\tau_p y^3)\rangle^2}{\langle x^4\tau_e\rangle^2}\right)\right]. \qquad \ldots\ldots(65)$$

If we put $\tau_e = (3\pi^{1/2}\mu_0 m)/4ex$ and $\tau_p = (BT^4y^2)^{-1}$, then we find

$$\langle x^8\tau_e\rangle/\langle x^4\tau_e\rangle = 6, \qquad \langle x^6\tau_e\rangle/\langle x^4\tau_e\rangle = 2,$$
$$\frac{\langle x^6\tau_e\rangle\langle x\tau_e(\tau_p y^3)\rangle}{\langle x^4\tau_e\rangle^2} = \frac{\pi^{1/2}}{2BT^4}, \qquad \frac{\langle x^3\tau_e(\tau_p y^3)\rangle}{\langle x^4\tau_e\rangle} = \frac{3\pi^{1/2}}{8BT^4},$$
$$\frac{\langle x\tau_e(\tau_p y^3)\rangle}{\langle x^4\tau_e\rangle} = \frac{\pi^{1/2}}{4BT^4}, \qquad \frac{\langle x^{-2}\tau_e(\tau_p y^3)^2\rangle}{\langle x^4\tau_e\rangle} = \frac{1}{4(BT^4)^2}.$$

Then

$$\frac{\kappa_e+\kappa_p{}'}{\sigma T} = \left(\frac{k}{e}\right)^2\left[2-\left(\frac{eQ_p}{k}\right)+\left(\frac{eQ_p}{k}\right)^2\left(\frac{4}{\pi}-1\right)\right]. \qquad \ldots\ldots(66)$$

Now, (eQ_p/k) can have quite large values at low temperatures, giving WFL numbers very different from the value of two given by the elementary theory. Thus, in n-germanium at 80° K (eQ_p/k) is 12·5, so that the WFL number must be about $32(k/e)^2$, i.e. 16 times the normal free-electron value. However, the thermal conductivity to which this corresponds is still negligible compared with κ_p in the germanium case. The form of (66) suggests that there will be a minimum value of the WFL number; this turns out to be $1{\cdot}06(k/e)^2$, when eQ_p/k is 1·82 and Q_p is $1{\cdot}6 \times 10^{-4}\,\text{v}\,\text{deg}^{-1}$. This will occur at between 150 and 200°K. The data used here are from Geballe and Hull (1954).

The effect of increasing the electron concentration will be threefold. Firstly, the electronic heat conductivity will increase through the increase in n directly. Secondly, saturation effects will diminish the terms which depend on ratios of the type $\langle x^4\tau_e(\tau_p y^3)\rangle/\langle x^4\tau_e\rangle$. Finally, we shall get some reduction in κ_p due to scattering of the phonons by the electrons.

§ 7. CONCLUSIONS

It will be seen that the effects of phonon disequilibrium on electron transport phenomena, and of electron disequilibrium on phonon transport phenomena are widespread. They are most noticeable in the realm of thermoelectricity, but we have shown that, under favourable conditions, they may be important both in electrical and thermal conduction. The mathematical difficulties are considerable, except in the case of a dilute semiconductor, but the effects of increasing carrier concentration can be seen qualitatively without difficulty.

ter Haar and Neaves have recently (1956) published work covering much the same ground as this paper. Their work unfortunately suffers from a fundamental error in the formulation of the collision integral ($\mathscr{L}_4(\Psi)$ in this paper) for the scattering of phonons by electrons. If the collisions are elastic for the electrons, which is the case at all but the lowest temperatures, then we must have the condition that a phonon of wave vector $\mathbf{q}$ cannot interact with electrons whose wave vectors are less than $\mathbf{q}/2$; this lower limit is ignored by ter Haar and Neaves. In consequence, they obtain the wrong variation of τ_{pe} with $\mathbf{q}$. Using the notation of the present paper, they find that $\tau_{pe}^{-1} \propto y$, whereas, using the correct lower limit, one obtains $\tau_{pe}^{-1} \propto y\exp(-y^2)$. This error results in far too much weight being given to the high-energy lattice vibrations and, in consequence, their expression for the lattice contribution to the Peltier coefficient is much too large and in disagreement with experiment.

Another field in which dragging effects could be important is in intrinsic semiconductors, in particular in the realm of thermal conduction. Here we have three interacting particle currents of electrons, holes and phonons, moving together, and it may certainly be anticipated that their behaviour will not be the same as if they were independent. The author hopes to describe these effects in detail in a subsequent paper.

ACKNOWLEDGMENTS

I wish to thank Dr. T. E. Allibone, the Director of the Research Laboratory, Associated Electrical Industries, Ltd., for permission to publish this paper.

REFERENCES

BOND, W. L., MASON, W. P., MCSKIMIN, H. J., OLSEN, K. M., and TEAL, G. K., 1950, *Phys. Rev.*, **78**, 176.
BROOKS, H., 1951, *Phys. Rev.*, **83**, 879 (A).
CONWELL, E., and WEISSKOPF, V. F., 1950, *Phys. Rev.*, **77**, 388.
DEBYE, P. P., and CONWELL, E., 1954, *Phys. Rev.*, **93**, 693.

Erginsoy, C., 1950, *Phys. Rev.*, **79**, 1013.
Frederikse, H. P. R., 1953, *Phys. Rev.*, **92**, 248.
Geballe, T. H., and Hull, G. W., 1954, *Phys. Rev.*, **94**, 1134; 1955, *Ibid.*, **98**, 940.
Gurevich, L., 1945, *J. Phys., Moscow*, **9**, 477.
ter Haar, D., and Neaves, A., 1955, *Proc. Roy. Soc.* A, **228**, 568; 1956, *Advanc. Phys.* (*Phil. Mag.* Suppl.), **5**, 241.
Herring, C., 1954 a, *Phys. Rev.*, **95**, 954; 1954 b, *Ibid.*, **96**, 1163.
Makinson, R. E. B., 1938, *Proc. Camb. Phil. Soc.*, **34**, 474.
McSkimin, H. J. M., Bond, W. L., Buehler, E., and Teal, G. K., 1951, *Phys. Rev.*, **83**, 1080.
Morin, F. J., 1954, *Phys. Rev.*, **93**, 62.
Morin, F. J., and Maita, J. P., 1954, *Phys. Rev.*, **96**, 28.
Parrott, J. E., 1954, *Proc. Phys. Soc.* B, **67**, 587; 1956, *Ph.D. Thesis*, University of Reading.
Sommerfeld, A., and Bethe, H., 1933, *Handbuch der Physik*, **24**/2 (Berlin : Springer).
Sondheimer, E. H., 1956, *Proc. Roy. Soc.* A, **234**, 391.
Wilson, A. H., 1953, *Theory of Metals*, 2nd Edn (Cambridge : University Press).

The Effect of Thin Films of Gold on the Mechanical Properties of Cadmium Single Crystals

By F. R. LIPSETT † and R. KING
Davy Faraday Research Laboratory, The Royal Institution, London

MS. received 25th February 1957

Abstract. Thin films of oxides, hydroxides or metals have been shown to increase the stress at which plastic deformation starts by an amount which cannot be explained by the strength of the films themselves. Most experiments have been on creep, showing the effect of oxides and hydroxides. This paper records an investigation into the effect of thin films of gold on the stress–strain curves of cadmium crystals.

The crystals were grown in an Andrade–Roscoe furnace and extended in a rigid frame machine with a Polanyi beam and photographic recording. The gold films were deposited by sputtering.

It was found that thin gold films increased the stress at which plastic deformation started by about 6 g wt mm^{-2}. The increase was independent of film thickness for films of 1500–24 000 Å. The nature of the effect of the film upon the stress–strain curves depended upon the orientation of the crystals.

§ 1. Introduction

One aspect of the deformation of metal single crystals which received little attention until the last few years was the effect of the surface condition on the mechanical properties. In 1936 Roscoe found that the resolved shear stress at the start of plastic deformation (critical shear stress) of cadmium single crystals was increased nearly 2½ times if a thin film of oxide was present on the surface, an increase too large to be accounted for by the strength of the film itself. The importance of surface condition in creep behaviour was brought to the fore by Andrade and Randall (1948), who showed that surrounding a creeping cadmium crystal with an electrolyte led to a large increase in the rate of flow of the crystals, due to the removal by the solutions of a thin coat of oxide which was hindering creep. The same effect was shown to take place with zinc crystals immersed in electrolytes (Andrade and Randall 1952) or acids (Harper and Cottrell 1950, Coffin and Weiman 1953), but it does not occur with polycrystalline zinc or cadmium (Andrade and Randall 1952, Pickus and Parker 1951, Phillips and Thompson 1950). The removal of thin hydroxide films produces a similar increase in creep rate (Phillips and Thompson 1950). On the other hand the addition of thin films of the metal copper have been shown to reduce the creep rate of zinc single crystals to a marked extent (Pickus and Parker 1951), even though the copper diffuses rapidly into the zinc.

The increase in the critical shear stress of cadmium single crystals shown by Roscoe (1936) was confirmed (Cottrell and Gibbons 1948, Makin 1952, Likhtman and Ostrovskiĭ 1953), and shown to occur also in crystals of zinc (Harper and

---

† Now at Division of Radio and Electrical Engineering, National Research Council, Ottawa 2, Ontario, Canada.

Cottrell 1950) and silver (Andrade and Henderson 1951). The addition of electroplated films of copper on to zinc, tin and lead crystals (Gilman and Read 1952) was found in each case to increase the critical shear stress. The critical shear stress for twinning is similarly increased (Gilman and Read 1952, King 1952). These experiments have allowed a qualitative explanation of the influence of surface condition on the plastic deformation of metal single crystals to be given in terms of the dislocation theory. It has been suggested by Frank (1949) that crystals can grow in reasonable times only if they contain dislocations. During deformation these dislocations may multiply (Frank and Read 1950) and move out through the surface of the crystal to produce slip lines. If a coating is present on the surface, however, the motion of the dislocations is impeded and they will tend to pile up behind the barrier of the coating until the stress reaches a value sufficient to enable them to break through. The sudden increase in creep rate when surface coatings are removed may then be accounted for, since in such a case the removal of the barrier to the motion of the dislocations allows them to escape through the surface more easily than before. This explanation is strengthened by experiments showing an abnormal after-effect in metal wires and crystals in torsion (Barrett 1953).

On the other hand it is not yet possible to give a detailed quantitative explanation of the role of surface condition in plastic deformation. Furthermore, certain features of the deformation in the presence of a surface coating, such as the production of slip lines and the effect on the stress–strain curve, have not been adequately described except in a very few cases (Gilman and Read 1952, Makin 1952). Experiments on the stress–strain behaviour of crystals are therefore essential in order to make possible a better understanding of the deformation process. The present work was undertaken in order, at least in some small part, to remedy this situation. In particular we have investigated the effect of thin films of the metal gold, which is soluble in cadmium, upon the stress–strain curves and production of slip lines in cadmium single crystals.

§ 2. EXPERIMENTAL DETAILS

2.1. *Outline of Experiments*

In experiments of the type described, the degree of reproducibility of behaviour from one crystal specimen to another is naturally of considerable moment. There are numerous factors which affect the measured critical shear stress and the shape of the stress–strain curve, and reproducibility of behaviour is lost unless attention is paid to these factors. In particular the purity of the crystals, and the geometry of the specimen and testing apparatus have important effects. The rate of extension also affects the value of the critical shear stress (Andrade and Randall 1952, Makin 1952), as well as the temperature of testing (Boas and Schmid 1929) and the diameter of the specimen (Makin 1952, Likhtman and Ostrovskio 1954). In the present experiments careful consideration was given to all these factors affecting the stress–strain curves, and the following test conditions were adopted.

All quantitative estimates of the effects of surface coatings were made between lengths cut from one single crystal, or, where this proved impracticable, from crystals of identical orientation grown by a seeding technique. The gauge length of the specimens, 28·8 mm, was the same for all tests. In this way variations due to purity and conditions of growth were eliminated. The diameter of the crystals used was 1·0 mm, which was greater than that at which Makin (1952) observed variations with the specimen diameter. The rate of extension was the same for all tests, and was set at 0·064 mm min^{-1}, or 0·22% per minute. In order to avoid errors due to inaccuracies in the determination of crystalline orientation and calibration of the extension apparatus, the experiments were confined to crystals in which the angle χ_0 initially contained between the basal plane and the specimen was 13–51°.

2.2. *Preparation and Examination of Specimens*

All crystals used in these experiments were grown from spectroscopically pure cadmium which had been melted and poured *in vacuo* in order to eliminate gaseous and oxide impurities. The cadmium billets obtained by vacuum casting were extruded into wire of 1·0 mm diameter in a small press. Three samples of cadmium were used, the analyses for which are given in the table. Cadmium 1 was kindly supplied by Dr. M. J. Makin, cadmium 2 was prepared in our laboratory, and cadmium 3 was supplied by Messrs. Johnson, Matthey and Co. Ltd. Cadmiums 1 and 2 were prepared from spectroscopically pure billets supplied by the latter firm.

Spectrographic Analyses of Samples of Cadmium

Cadmium	Metallic impurities (%)						Crit. Shear Stress
No.	Pb	Cu	Th	Ag	Ca	Fe	(g wt mm^{-2})
1	Trace		Trace				20·0
2		Trace		Trace	Trace	Trace	29·0
3	<0·00005				<0·00005		24·5

In order to transform the cadmium wire into single crystals, lengths of wire were first cleaned carefully with Silvo and benzene and rolled straight between pieces of plate glass separated by clean sheets of paper. The length of wire was then placed in a Pyrex quill which had been specially selected to have a bore of 1·2–1·7 mm, and which had been cleaned with chromic acid and distilled water and outgassed by heating in a vacuum system. After insertion of the wire the quill was re-sealed and evacuated, and after a suitable degassing period argon was admitted to the quill. It was then sealed off, with the cadmium wire ready for growth as a single crystal. The single crystals were grown in a furnace of the Andrade–Roscoe (1937) type with water-cooled ends. Power for the furnace was supplied by a battery of accumulators, which ensured a uniform furnace temperature during growth. The furnace was driven over the Pyrex container on kinematically designed track at the rate of 1·44 in. hr^{-1}.

After growth the crystals were carefully removed from their containers for x-ray examination. They were placed in a milled groove in a long slotted bar which was arranged to travel accurately at right angles to a beam of x-rays. A camera was placed in a kinematic mounting between the x-ray tube and crystal in such a position that the distance between the film and the specimen was 3·0 cm. Back-reflection photographs were taken at intervals along the length of the crystal, and, since the orientation of the crystal relative to the x-ray beam and camera was not changed for successive photographs, it was possible to detect visually without detailed analysis of the photographs whether the wire retained the same crystalline orientation throughout its length. This could be done merely by superimposing the developed photographs.

The X-ray photographs were examined for any signs of bi-crystalline structure or non-uniformity of orientation. If these were present the crystal was discarded. If the photographs showed no sign of abnormality a Greninger chart was employed to prepare a stereographic projection of the crystal lattice and the angles χ_0 between the basal plane (0001) and the specimen axis, and λ_0 between the glide direction [1120] and the specimen axis were determined. The accuracy of this determination was estimated to be to ±1°. All specimens were also examined visually for any signs of distortion or irregularity. After examination the long crystals were cut into sections and each section was sealed off *in vacuo* in Pyrex tubing for storage.

When a crystal of a certain orientation was desired, in order to allow a large number of tests to be carried out on crystals of identical orientations, a technique was developed to grow crystals from a 'seed' crystal of the desired orientation. A 'seed' about 4 in. long and a length of wire were sealed off in an argon atmosphere in selected Pyrex tubing in the usual manner. The Pyrex container was then tilted slightly from the horizontal so that the end of the seed was brought into contact with the wire. The junction of the wire and seed was heated very gently with a flame until the cadmium melted and a weld was formed. Welding in an argon atmosphere in this manner prevented the formation of oxides in the weld, which tend to prevent the satisfactory growth of the crystal from the seed. The container was now ready to be placed in the Andrade–Roscoe furnace. The furnace was raised to the appropriate temperature and placed in such a position over the container that the molten section of the cadmium was in the middle of the seed crystal. When the furnace was set in motion, the growth started from this point of the seed, and crystals of predetermined orientation could easily be grown. This method of seeding was so successful that no difficulty was found in obtaining sufficient test lengths for any comparisons, it always being possible to produce additional lengths of a desired orientation.

2.3. *Deposition of Gold Films*

The method of sputtering was selected for the deposition of the films, because of the high uniformity of film thickness obtainable and the ease of controlling the thickness. In addition to these advantages, the setting up of internal stresses in the crystal surface, sometimes encountered with electrolytic deposition, was avoided. A series of experiments was carried out in order to allow the design of a sputtering apparatus suitable for the present experiments, which has been described elsewhere (Lipsett 1956). The sputtered films were found to be of remarkably high uniformity both around the circumference and along the length of the specimens.

A diagram of the sputtering chamber used in the present experiments is given in figure 1. The chamber rested on a brass table placed in a large vacuum system, and the discharge took place between a gold-foil cathode and steel anodes at either end of the chamber. The crystal and its supporting piece were electrically insulated from the top anode. When a specimen was to be sputtered before extension a special procedure was adopted in order to avoid damage during insertion and removal from the sputtering chamber. A supporting wire of cadmium

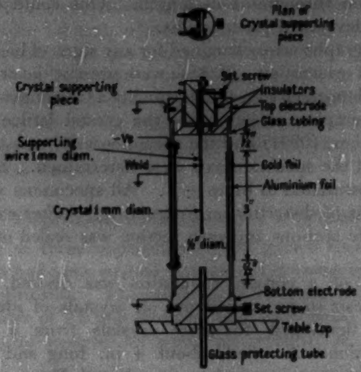

Plan of Crystal supporting piece
Set screw
Crystal supporting piece
Insulators
Top electrode
Glass tubing
Supporting wire 1mm diam.
-Vs
Weld
Gold foil
Aluminium foil
Crystal 1 mm diam.
3"
½" diam.
Bottom electrode
Set screw
Table top
Glass protecting tube

Figure 1. Diagram of the sputtering chamber used for depositing gold films on single-crystal specimens.

was first welded to the specimen. In order to do this the specimen and wire were placed together in a milled groove in a long piece of Syndanyo board, with the junction between wire and crystal over a ⅛ in. diameter hole. After a couple of practice runs on pieces of wire a tiny gas–oxygen flame was applied to the junction of wire and crystal until they melted and flowed together in a smooth weld. The practice runs were essential in order to ensure that the flame was adjusted to the right heat and that the timing of the welder was correct.

After welding, the wire and crystal were carefully inserted into a glass protecting tube. The tube was placed in a hole in the fitment labelled crystal supporting piece in figure 1. This fitted into the top anode of the sputtering chamber. The glass tube was secured by a set-screw, and the end of the supporting wire was gently tapped out of the tube to rest between a flat surface on the supporting piece and a spring-loaded plate. The plate was tightened up to secure the wire at such a height that the crystal hung in the centre of the sputtering tube. The crystal supporting assembly was now placed in the electrode, with the glass tube projecting through the bottom electrode. The glass tube was released by the set-screw and lowered through the bottom electrode to leave the crystal exposed and ready for sputtering.

After insertion of the specimen the vacuum system was outgassed at a pressure of less than 0·1 micron for at least an hour. Sputtering was then carried out at a pressure of about 10 microns, with a discharge of 350 volts. The current under these conditions was about 10 mA, and the temperature of the specimen 55°C. Films of 1500–24 000 Å were deposited on various specimens.

Upon the completion of sputtering the entire top electrode was removed from the chamber and supported in a clamp. The supporting wire was cut off a short distance above the weld, and the specimen was placed in the rigid frame machine for extension without disturbance of the sputtered film. The section of the crystal adjacent to the weld was not included in the gauge length for extension.

2.4. *Extension Apparatus*

For extending the single crystal specimens an apparatus of rigid frame type was employed, the stress being measured by means of a Polanyi beam and a permanent record of load against extension produced on photographic paper. The specimen was extended between two grips. One of the grips was rigidly attached to the base of the machine, while the other formed part of a stirrup resting on the centre of the Polanyi beam. The beam rested in a carriage which was drawn upwards by a machined screw driven by a slow-speed motor. The load was applied to the specimen through the Polanyi beam, so that the deflection of the beam was a measure of the load on the specimen.

The deflection of the Polanyi beam was measured by an optical method. At each end of the beam was fixed a surface-aluminized mirror. Light from a Pointolite lamp was focused on a 1/64 in. diameter ball bearing, which was glued to the end of a blackened gramophone needle. The image of the Pointolite source produced by the surface of the ball bearing acted as a point source of great intensity. The light was focused by a lens, after deflection by the mirrors on the Polanyi beam, on to photographic paper attached to the surface of a rotating drum. The drum was placed six feet from the machine, and at this distance the magnification of the Polanyi beam deflection was 400. To provide a reference line on the photographic paper a spot of light from a small lamp was focused on the surface, and as the drum rotated a horizontal line was printed. The height of the drum was such that traces showing up to 15 in. deflection on the photographic paper could be obtained. The Polanyi beams used were calibrated by attaching small weights to the stirrup and recording the deflection of the light spot on a scale attached to the drum. The deflection was linearly dependent on the weight applied to the beam.

The load–extension data obtained from the photographic traces were converted into the conventional stress–glide curves by use of the following relations (Schmid and Boas 1950):

$$\text{Resolved shear stress} = \frac{F\sin\chi_0}{A_0\, l/l_c}\left\{\left(\frac{l}{l_c}\right)^2 - \sin^2\lambda_0\right\}^{1/2} \qquad \ldots\ldots(1)$$

$$\text{Glide strain} = \frac{1}{\sin\chi_0}\left[\left\{\left(\frac{l}{l_c}\right)^2 - \sin^2\lambda_0\right\}^{1/2} - \cos\lambda_0\right] \qquad \ldots\ldots(2)$$

where F = load in g wt, l = extended length of specimen, l_c = effective gauge length, and A_0 = initial cross-sectional area of specimen. The effective gauge length, or length of the specimen over which the glide planes are not locked by the action of the grips, is defined by $l_c = L - d\cot\chi_0$, L = gauge length = 28·8 mm, d = specimen diameter $\simeq$ 1·0 mm.

When converting the load–extension data a small correction was made for the deflection of the Polanyi beam in order to determine the extension of the specimen. The diameter of the specimen was measured every 5 mm along the gauge length of the specimen by means of a microscope with a micrometer eyepiece, and the diameter thus obtained was used in calculating the initial cross-sectional area of the specimen. In all specimens the diameter was very uniform, varying by less than 3%, though it was found in some cases that the cross section was slightly elliptical.

§ 3. RESULTS

3.1. *Outline of Experiments*

Tests were first carried out to examine whether annealing treatments altered the properties of the single crystals. It was found that annealing after growth before mechanical testing produced no significant effect on the critical shear stress. Results with all three samples of cadmium were reproducible within reasonable limits. Crystals of a number of different orientations, ranging from $\chi_0 = 13°$ to $\chi_0 = 51°$ were examined for the effects of sputtered films. Three particular orientations (low, intermediate and high values of χ_0) were examined in detail, sufficient samples of identical orientation being produced by seeding. The effects of various thicknesses of sputtered film were observed and certain general conclusions were reached.

3.2. *Effect of Films on the Critical Shear Stress for Slip*

It was found that in all tests the presence of a gold film increased the critical resolved shear stress. There was considerable scatter in the results obtained but it was possible to make the following observations from a number of different tests.

Films of thickness ranging from 1500–24 000 Å were coated on crystals ranging in orientation from $\chi_0 = 13°$ to $\chi_0 = 51°$. The average effect of a film of given thickness on a crystal of given orientation was found by subtracting the average value of the critical shear stress for the tests on the uncoated crystals—usually as grown and annealed crystals—from the average value of the critical shear stress for the coated specimens. When this procedure was followed it was found that the average effect was the same for all thicknesses of film in the range tested and was also independent of the orientations of the crystals. Crystals made from the three different samples of cadmium wire differed somewhat in their critical shear stresses (see table), but the effect of the film was the same in all three cases. The critical shear stress was raised on an average in all cases by 6 g wt mm^{-2}. This was about 25% of the critical shear stress of the uncoated crystals.

The constancy of the effect was very surprising. It would be expected that the effect would fall along some smooth curve as the thickness of the film was reduced, but tests with thinner films showed such variation in behaviour that no conclusions could be drawn.

3.3. *Effect of Films on the Stress–Strain Curve*

The effects of the presence of the films on the stress–strain curves were twofold. The influence of the film upon the stress–strain curve was dependent upon the orientation of the crystal. The first most noticeable effect was that the stress–strain curves of the sputtered specimens showed irregularities in slope, humps being recorded in most cases. Irregularities were observed in uncoated crystals, particularly those of high angle χ_0, but the irregularities were more marked in the sputtered specimens. The second effect was in the general trend of the stress–strain curve as extension proceeded. This was systematically dependent upon crystalline orientation and is typified in the following examples.

3.3.1. *Crystals with small values of χ_0.*

Figure 2 shows stress–glide curves for specimens of a crystal of cadmium 1 of orientation $\chi_0 = 14{\cdot}5°$ and $\lambda_0 = 15°$. In view of the minor irregularities and small differences between sections of the same crystal the two curves shown have been obtained by averaging. The feature which is of interest is that the stress–strain curve for the coated specimens starting at 6 g wt mm^{-2} above that of the uncoated specimens continues at about the same height above that curve right to the rapid increase in work hardening which begins at about 200% glide. This behaviour was typical of crystals for which the angle χ_0 was small, *circa* 15°.

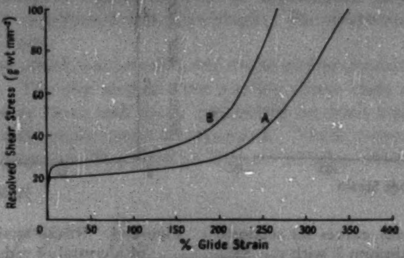


Figure 2. Stress–strain curves for sections of a crystal of cadmium 1, with $\chi_0=14{\cdot}5°$, $\lambda_0=15°$. A, as grown and annealed sections; B, sections with gold films 12 000 Å thick.

3.3.2. *Crystals with large values of χ_0.*

It was typical of crystals in which the value of χ_0 was large that the load–extension curves showed an apparent softening of the crystal after the first yield had taken place. This phenomenon may be attributed to the geometrical consequences of testing a single crystal in tension under conditions in which the two ends are constrained to move in a given line. The consequences of the 'geometrical softening' were pointed out by Andrade and Roscoe (1937) and are discussed in some detail by Schmid and Boas (1950, p. 118). With the short specimens used in the experiments here recorded it was found that the calculated stress–glide curves showed some initial softening in the case of relatively large values of χ_0. This was probably due, in part at least, to the inadequate correction for the locking effect at the grips. This feature is shown in figure 3.

The uncoated crystals showed a rapid softening after the start of plastic deformation, while the sputtered sections did not show as much softening and this was spread over a greater amount of the extension of the specimen. The softening appeared to be 'rounded off' in the sputtered specimens. In contrast to the case for the low-angle crystals, the strengthening effect is not maintained at large extensions. At 150% glide the resolved shear stress is approximately the same for coated and uncoated crystals. This behaviour is in many respects reminiscent of the behaviour of crystals coated with hydroxide layers (Makin 1952).

3.3.3. *Crystals with intermediate values of χ_0.*

The effects of coatings on crystal specimens of intermediate values of χ_0 were found to be qualitatively what would be expected from the observations

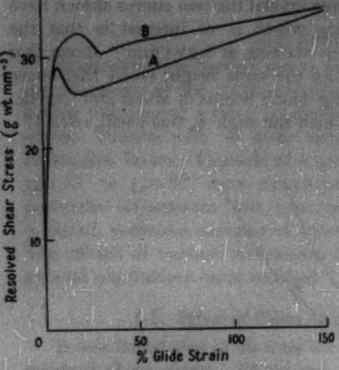


Figure 3. Stress–strain curves for sections of a crystal of cadmium 3, with $\chi_0=51°$, $\lambda_0=51°$. A, as grown and annealed sections; B, sections with gold films 9000–24 000 Å thick.

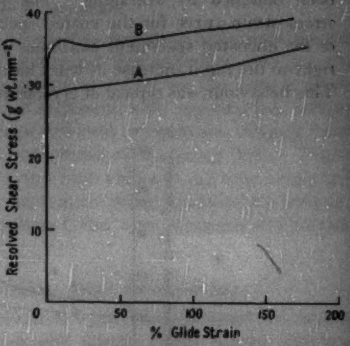


Figure 4. Stress–strain curves for sections of a crystal of cadmium 2, with $\chi_0=31°$, $\lambda_0=32°$. A, as grown sections; B, sections with gold films 12 000 Å thick.

on crystals with high and low values of χ_0. Figure 4 is typical of the behaviour of intermediate orientations. There was no marked geometrical softening, but as the extension proceeded the strengthening effect appeared to diminish and the two curves approached each other with increasing glide.

3.3.4. *The effect of the rate of extension.*

In view of the rather surprisingly constant value of the effect of gold films on the critical shear stress, a few experiments were carried out to determine whether the effect was appreciably dependent upon the rate of straining of the specimen. Specimens were extended at four times the normal rate of testing, that is at $0{\cdot}25\ \mathrm{mm\ min^{-1}}$. While, as was expected, the critical shear stress of both uncoated and coated crystals was raised by increasing the rate of extension, there was no significant increase in the effect of the film.

3.4. *Observations on the Specimens during the Process of Deformation*

Many crystals were examined at various stages during the deformation and after final removal from the extension apparatus in a search for any visual differences between the coated and uncoated specimens.

In both types of specimen lack of uniformity of extension along the length of the specimen was observed. This was particularly true of high-angle crystals in the untreated condition but was more marked for the sputtered specimens in all cases. The lack of uniformity took the form of slip starting initially in one or more sections of the crystal and gradually extending along the length of the crystal. At intervals slip would start in another region and extension would also spread from this region until at relatively large elongations the slipped regions would meet, after which the rate of extension appeared to be uniform along the length of the crystal. As far as could be judged the irregularities observed were in keeping with the irregularities in the stress–strain curve.

A very careful search was made for any difference in the appearance of the slip bands themselves between coated and uncoated specimens. A specimen was coated to a thickness of 1500 Å over only half its length by protecting one half while it was in the sputtering chamber by means of a fairly closely fitting glass tube. This produced a relatively sharp step between the sputtered and unsputtered portions. The specimen was then extended 9% and the two portions examined for comparison. The strengthening effect of the film was made quite apparent. The uncoated region showed an appreciable density of slip lines relatively uniformly distributed along its length; the sputtered portion showed a few isolated slip lines.

One observation which was consistently made and promotes some speculation was the following. The slip steps on the surface where edge dislocations broke through the film were clean and sharp and bright on both coated and uncoated specimens and no differences could be observed. These correspond to points A in figure 5. The steps are illustrated in figure 6 (Plate), which is a photograph of a slightly extended crystal with a film of 1500 Å thickness. At the sides of the crystal, however, where screw dislocations would break through the surface (points B of figure 5), differences were observed. In an uncoated crystal slip lines disappear at points corresponding to points B in figure 5, since at this point no step is formed in the surface. In coated crystals, however, the evidence of the existence of a slip line was very marked at these points. At relatively small extensions the surface shows corrugations parallel to the slip direction and as extension proceeds the corrugations increase in roughness until the film or at least an appreciable thickness of it ceases to adhere to the crystal and begins to flake off at the surface. An early stage in this process is shown in figure 7 (Plate), in which the direction of view lies along the slip planes. Figure 8 (Plate) shows a late stage in the breaking away of the film, the direction of view being roughly normal to the slip planes. From these observations it is clear that, as is to be expected, the film is not highly resistant to shear forces normal to it, but resists to a much higher degree the shear forces lying in its plane. Or in other words the passage of the edge dislocations through the film is achieved more readily than the passage of the screw dislocations. The film prefers to tear away from the crystal rather than allow the passage of the screw dislocations, whereas there was never any evidence of the film losing its adherence where the edge dislocations broke through, even after considerable extension during which the film had flaked appreciably at the sides.

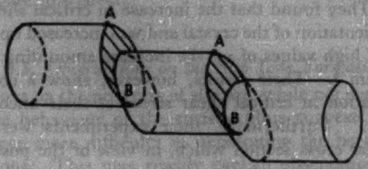


Figure 5. Diagram illustrating translational slip in a cadmium crystal. The gold films were easily broken at points A, and flaking off commenced at points B.

One other observation on the flaking of the films was of interest. When the films flaked the surface was never left clean and bright except at the new surface formed by slip steps. This, while inconclusive, argued that the whole thickness of the film did not flake off but that there was a thin adherent layer which was retained.

§ 4. DISCUSSION

4.1. *The Constancy of the Critical Shear Stress Effect*

The observations suggest that it is only a certain limiting layer which is really effective in increasing the critical shear stress. This is not unreasonable. Gold is soluble in solid cadmium and it is likely that there will be formed a very thin layer of solid solution at the surface. The lattice strains in this layer would present an effective obstacle to the passage of dislocations, while the pure gold outside this region would be expected to deform or rupture readily once the stress necessary to drive the dislocations through this layer had been achieved.

The results are, however, in contrast with those reported by Gilman and Read (1952), who observed the effect of electrolytically deposited layers of copper on zinc crystals. They found that the increase in critical shear stress depended markedly on the orientation of the crystal and was increased up to twice its normal value in crystals of high values of χ_0, the increase amounting in such cases to as much as 23 g wt mm^{-2}. Their results, however, show a considerable scatter, there being a variation in critical shear stress of their uncoated crystals from 26 to 58 g wt mm^{-2}. Furthermore, their experiments were performed with crystals of triangular cross section which, in view of the possible differences in resistance of films to screw and edge dislocations, cannot strictly be compared with crystals of circular cross section.

The observations are qualitatively in accord with those of Gumbrell (1951), on the effects of oxides and hydroxide films. In both cases the effect of the film on the critical shear stress both for slip and twinning increased to a limiting value as the thickness of the film was increased.

4.2. *Stress–Strain Curves of Sputtered Specimens*

The variation with crystal orientation of the differences between the coated and uncoated crystals is not readily explained. One or two possibilities suggest themselves, however. If fine slip made any significant contribution to the deformation of the clean crystal, then since it would be expected to be absent in a coated crystal, the stress–strain curve for the latter should rise more rapidly than that for the uncoated crystal, and the difference between the two curves should increase with increasing extension. There was no evidence of this occurring to any significant extent, though lack of reproducibility of behaviour might certainly have masked a small effect, as would also the effect next to be considered.

If, after the initiation of a given slip step, further glide occurred on the same slip plane as the stress was increased, the amount of glide in any one plane once started being a function of the stress, this would mean that the stress–glide curve of the coated specimen should have a smaller slope than that of an uncoated specimen, so that the difference between the curves should decrease with increasing extension. The evidence on this point is certainly inconclusive, but it might be tentatively suggested as an explanation of the variation of the behaviour with crystal orientation. In crystals of small χ_0, the region of accommodation between the slipped region and the portion locked in the grips

—a region in which there must be curvature of the slip planes—would affect to some degree most, if not all, of the operative slip planes. Curvature of the planes would be expected to inhibit continued slip so that the stress–strain curve of the coated crystal would not be expected to approach that of the uncoated crystal. When the angle χ_0 is large, on the other hand, the rotation of the planes for a given extension is greater and the curvature in the region of accommodation is more intense, but the number of effective slip planes affected by the curvature would be expected to be appreciably less. Thus, if slip continued on the planes which were not affected by curvature the stress–strain curve of the coated crystal should approach that of the uncoated crystal with increasing extension.

Alternatively, it may be argued that in the region of accommodation at the ends in crystals of high χ_0 the curvature is more intense than in crystals of lower χ_0, and in these regions breaking away of the film is initiated which spreads its effect along the crystal. Thus in high-angle crystals the effect of the film is lost as extension proceeds.

Neither of these explanations is entirely satisfactory. More work is clearly needed to investigate this effect.

4.3. *Observations on the Slip Line Formation*

The observations on the progress of slip are, while not conclusive, of potential significance. The behaviour of the surface films suggests that the screw dislocations experience more difficulty in penetrating the barrier of the film than do edge dislocations. This may readily explain observations which have been made, for example by Gilman and Read (1952), that the mechanical behaviour of a crystal depends on the shape of the cross section and of the orientation of the glide direction in that cross section.

ACKNOWLEDGMENTS

This work forms part of a thesis submitted to the University of London by one of us (F. L.) for the degree of Doctor of Philosophy. He is grateful to the Managers of the Royal Institution for the award of an Overseas Research Scholarship.

We wish to make grateful acknowledgment to Professor E. N. da C. Andrade, who suggested the investigations described, and to Dr. M. J. Makin, Atomic Energy Research Establishment, Harwell, for supplying some of the cadmium wire used and for helpful discussions.

REFERENCES

ANDRADE, E. N. DA C., and HENDERSON, C., 1951, *Phil. Trans. Roy. Soc.* A, **244**, 177.

ANDRADE, E. N. DA C., and RANDALL, R. F. Y., 1948, *Nature, Lond.*, **162**, 890; 1952, *Proc. Phys. Soc.* B, **65**, 445.

ANDRADE, E. N. DA C., and ROSCOE, R., 1937, *Proc. Phys. Soc.*, **49**, 152.

BARRETT, C. S., 1953, *Acta Metallurg.*, **1**, 2.

BOAS, W., and SCHMID, E., 1929, *Z. Phys.*, **57**, 575.

COFFIN, F. D., and WEIMAN, A. L., 1953, *J. Appl. Phys.*, **24**, 282.

COTTRELL, A. H., and GIBBONS, D. F., 1948, *Nature, Lond.*, **162**, 488.

FRANK, F. C., 1949, *Disc. Faraday Soc.*, **5**, 49.

FRANK, F. C., and READ, W. T., 1950, *Phys. Rev.*, **79**, 722.

Gilman, J. J., and Read, T. A., 1952, *J. Metals, N.Y.*, **8**, 875.
Gumbrell, S. M., 1951, *Ph.D. Thesis*, University of London.
Harper, S., and Cottrell, A. H., 1950, *Proc. Phys. Soc.* B, **63**, 331.
King, R., 1952, *Nature, Lond.*, **169**, 543.
Likhtman, V. I., and Ostrovskiĭ, V. S., 1953, *Dokl. Akad. Nauk S.S.S.R.*, **93**, 105.
Lipsett, F. R., 1956, *Brit. J. Appl. Phys.*, **6**, 442.
Makin, M. J., 1952, *Ph.D. Thesis*, University of London.
Phillips, D. J., and Thompson, N., 1950, *Proc. Phys. Soc.* B, **63**, 839.
Pickus, M. R., and Parker, E. R., 1951, *J. Metals, N.Y.*, **3**, 792.
Roscoe, R., 1936, *Phil. Mag.*, **21**, 399.
Schmid, E., and Boas, W., 1950, *Plasticity of Crystals* (London : F. A. Hughes).

RESEARCH NOTES

Temperature Variation of Microwave Absorption Coefficient in Ethyl Chloride

By KRISHNAJI AND G. P. SRIVASTAVA
Department of Physics, University of Allahabad, India

MS. received 21st January 1957

The absorption of microwaves by a heavy molecule is mostly due to rotational transitions in this region. The microwave absorption coefficient due to a rotational line can be expressed (Van Vleck and Weisskopf 1945) as

$$\alpha=\frac{8\pi^2N}{3\boldsymbol{k}T}f_vf_r|\mu_{ij}|^2\nu^2\left[\frac{\Delta\nu}{\Delta\nu^2+(\nu-\nu_0)^2}+\frac{\Delta\nu}{\Delta\nu^2+(\nu+\nu_0)^2}\right]\qquad\ldots\ldots(1)$$

where ν_0 (cm^{-1}) is the resonant frequency, $\Delta\nu$ (cm^{-1}) is the pressure broadening parameter, ν (cm^{-1}) is the operating frequency and μ_{ij} is the dipole moment matrix element for the given transition. The factors f_v and f_r are the fractions of molecules in the particular vibrational and rotational states respectively. In the case of asymmetric molecules at sufficiently high temperature (Townes and Schawlow 1955)

$$f_r=(2J+1)\exp\left(-\frac{W_{ij}}{\boldsymbol{k}T}\right)\left[\frac{\pi}{ABC}\left(\frac{\boldsymbol{k}T}{h}\right)^3\right]^{-1/2}\qquad\ldots\ldots(2)$$

where A, B and C are rotational constants in cycles per second. Substituting the value of N (number of molecules per cm^3) $=9{\cdot}68\times10^{19}\,p_{\mathrm{cm}}/T$, and f_r in the expression (1) we get

$$\alpha=\frac{\text{(constant)}}{T^{7/2}}\left[\frac{\Delta\nu}{\Delta\nu^2+(\nu-\nu_0)^2}+\frac{\Delta\nu}{\Delta\nu^2+(\nu+\nu_0)^2}\right]\qquad\ldots\ldots(3)$$

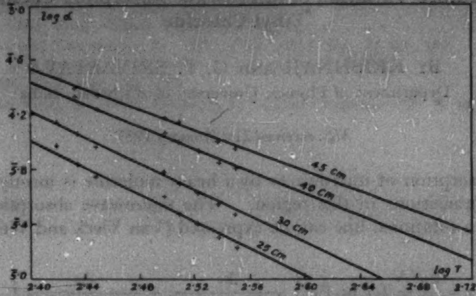
since $W_{ij}\ll\boldsymbol{k}T$ for most of the transitions. When ν is far removed from ν_0 and the pressure is sufficiently low that the inequality $\Delta\nu\ll(\nu-\nu_0)$ holds, the expression (3), relating α to temperature, reduces to

$$\alpha=\frac{\text{(constant)}}{T^{7/2}}\Delta\nu.\qquad\ldots\ldots(4)$$

In a previous communication (Krishnaji and Srivastava 1957) it has been shown that in ethyl chloride the microwave absorption in the 3 cm region is almost entirely due to low frequency Q branch rotational lines most of which are practically at zero wavenumber. The fraction of molecule taking part in these transitions is quite small. Since the absorption in this region due to rotational lines of P and R branches is negligible, the temperature dependence of α is represented by (4) to a fair degree of accuracy.

It is, therefore, apparent that the absorption coefficient should vary as $T^{-9/2}$ in support of Anderson (1952) or Margenau (1951) pressure broadening theory ($\Delta\nu\alpha T^{-1}$) and T^{-4} according to ‘ hard sphere ’ kinetic theory ($\Delta\nu\alpha T^{-1/2}$).

The microwave absorption coefficient for ethyl chloride has been measured at 8780 Mc/s by a fixed frequency spectrograph. Reflex klystron 2K39 has been used as the signal source and the absorption cell is a seven feet X-band waveguide, bent into a U-shaped piece so that it could be kept inside a constant temperature chamber. The measurements have been made at temperatures ranging from 262°K to 354·4°K and at a pressure ranging from 25 cm to 75 cm of mercury. The data are presented in the figure as graphs between log α and log T. The table gives the temperature dependence of α at different pressures as calculated from our experimental data.



Variation of log α with log T in ethyl chloride at different pressures.

Let $\alpha = (\text{constant}) T^{-x}$.

Variation of the Temperature Dependence of α with Pressure.

Pressure (cm Hg)	25	30	40	45
x	4·9	4·4	4·1	3·2

Taking into account the fluctuations in our measurements we find that α varies as $T^{-4\cdot9\pm0\cdot5}$ at a pressure of 25 cm of mercury. Within the range of experimental errors this temperature variation of α supports the view that the variation of $\Delta\nu$ as T^{-1} is more probable than its variation as $T^{-1/2}$.

The temperature dependence of the pressure broadening parameter $\Delta\nu$ will enable us to investigate better the inter-molecular forces for pressure broadening in the microwave spectra.

The authors are grateful to the Chairman, Radio Research Committee, Council for Scientific and Industrial Research, for permission to publish the results and the Council for financial assistance.

REFERENCES

ANDERSON, P. W., 1952, *Phys. Rev.*, **82**, 809 (L).
KRISHNAJI, and SRIVASTAVA, G. P., 1957, *Phys. Rev.*, in the press.
MARGENAU, H., 1951, *Phys. Rev.*, **82**, 156.
TOWNES, C. H., and SCHAWLOW, A. L., 1955, *Microwave Spectroscopy* (New York: McGraw-Hill).
VAN VLECK, J. H., and WEISSKOPF, V. F., 1945, *Rev. Mod. Phys.*, **17**, 227.

The Effect of Grain Size on the Magnetothermal Properties of Ferrites

By D. A. CHRISTOFFEL†

Department of Physics, University of Nottingham

Communicated by L. F. Bates; MS. received 15th March 1957

At a recent conference on ferrites Bates and Christoffel (1957) described experiments carried out on the magnetothermal properties of a number of ferrites. The purpose of this note is to give some further data about the behaviour of the ferrite $NiFe_2O_4$ and some data about a nickel ferrite in which cobalt is substituted to give a composition which approximates to $Ni_{0.98}Co_{0.02}Fe_2O_4$.

In figures 1 and 2 are given the thermomagnetic curves for $NiFe_2O_4$ specimens with particle sizes 1–5 μ and 5–10 μ, respectively. These specimens were kindly supplied by Dr. A. W. Simpson of the Plessey Company Ltd. The usual procedure of Bates and his collaborators was followed. Starting with the specimen in a field of 600 Oe, it was demagnetized step by step, and the adiabatic temperature changes were recorded as described in the paper submitted at the Conference on Ferrites which will shortly be published in the *Proceedings of the Institution of Electrical Engineers*. The lowest curve in figure 1 represents the theoretical changes in temperature which one would obtain if the magnetocaloric effect of Weiss and Forrer alone were present. The curve next above the lowest gives the experimental measurements. At the positive coercive point the curve splits into two portions, that labelled Q_T' being the actual experimental results obtained as already described, while Q_R' gives the points which are obtained when small backward and forward steps are made along the magnetization cycle as described by Bates and Sherry (1955). The curve Q_R'' is obtained by taking the difference between the curve Q_R' and the lowest curve, while that

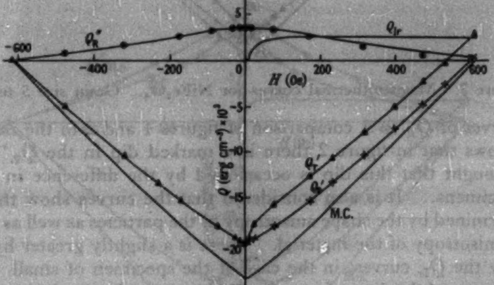


Figure 1. Thermomagnetic curves for $NiFe_2O_4$. Grain size 1 to 5μ.

---

† Now at 15, Lismore Street, Fitzroy, New Plymouth, New Zealand.

labelled Q_{Ir} is obtained from the difference between Q_T' and Q_R'. Similar remarks apply to figure 2.

On the Q_R'' curve have been entered certain points which have been calculated on the assumption that rotation processes alone are responsible for the magnetic changes observed in these ferrites over and above those due to the magnetocaloric effect. The factor b, which occurs in the expression $b\int H dI_R$ of the Stoner and Rhodes theory (1949), was assumed equal to $-0{\cdot}32$ in the case of figure 1 and to $-0{\cdot}16$ in the case of figure 2, in order to obtain best fit with the curves of Q_R''. A comparison of figures 1 and 2 in the neighbourhood of $H=0$ shows that in figure 2 there is a marked dip in the Q_R'' curve. It is therefore thought that this dip is occasioned by the difference in grain size in the two specimens. It is also considered that the curves show that the values of b are determined by the shape anisotropy of the particles as well as the magnetocrystalline anisotropy of the material. There is a slightly greater hysteresis loss, as shown by the Q_{Ir} curves, in the case of the specimen of small particle size.

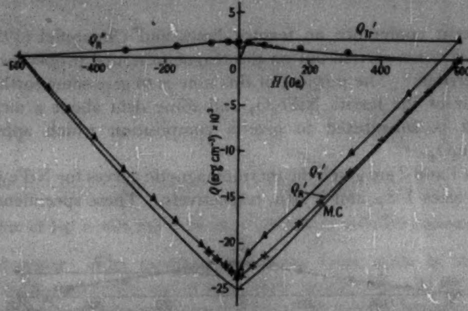


Figure 2. Magnetothermal curves for $NiFe_2O_4$. Grain size 5 to 10μ.

The presence of the dip in figure 2 would suggest that in the case of the larger particles boundary movements as well as rotation of the magnetic vectors can occur in the particles. Evidence obtained from these and other ferrites would indicate therefore that there is probably a critical size below which boundary movements do not make an effective contribution to magnetic changes. In the case of the ferrites now under discussion this critical size is probably between 2 and 8μ. A similar result has been found by Guillaud and Paulus (1956) in the case of a manganese–zinc ferrite, for which the critical size was 5·6μ.

In figures 3 and 4 are given the results for the $Ni_{0{\cdot}98}Co_{0{\cdot}02}Fe_2O_4$ ferrite for particle sizes lying between 1–5μ and 5–10μ, respectively. In this case no dip is found in the Q_R'' curve. On the other hand it was impossible to reproduce the shapes of the Q_R'' curves by assuming constant values of the quantity b. In figures 3 and 4 the curves $b\int H dI_R$ have been plotted for values of $b=-1{\cdot}2$ and $-2{\cdot}4$; these values were estimated by an examination of the magnetization curves, following a method which it is hoped to publish later. The differences between the $b\int H dI_R$ and the Q_R'' curves in figures 3 and 4 are very marked. This cobalt substituted nickel ferrite is unusual in that it appears to possess uniaxial anisotropy properties superimposed upon those of cubic anisotropy

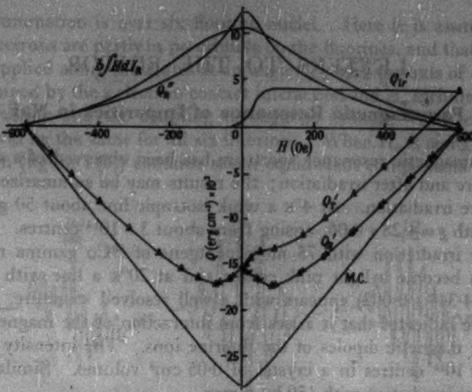


Figure 3. Thermomagnetic curves for $Ni_{0\cdot98}Co_{0\cdot02}Fe_2O_4$. Grain size 1 to 5μ.

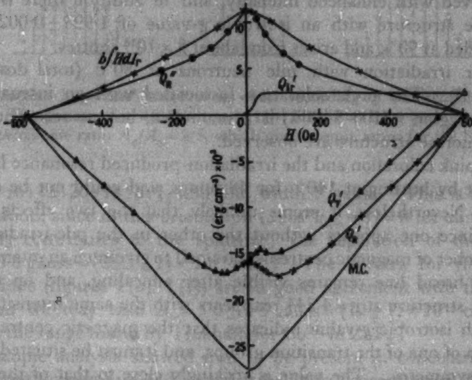


Figure 4. Thermomagnetic curves for $Ni_{0\cdot98}Co_{0\cdot02}Fe_2O_4$. Grain size 5 to 10μ.

normally found with the unsubstituted nickel ferrites, and there is here no evidence of special behaviour of particles of a critical size. The experiments are, however, to be continued at Nottingham on nickel ferrites with higher cobalt contents after subjecting the specimens to different field-temperature treatments.

I wish to thank Professor L. F. Bates for supervising this work and the New Zealand Defence Scientific Corps for a maintenance allowance.

REFERENCES

Bates, L. F., and Christoffel, D. A., 1957, *Proc. Instr. Elect. Engrs*, B, Suppl. 5, 231.
Bates, L. F., and Sherry, N. P. R., 1955, *Proc. Phys. Soc.* B, **68**, 642.
Guillaud, C., and Paulus, M., 1956, *C. R. Acad. Sci., Paris*, **242**, 2525.
Stoner, E. C., and Rhodes, P., 1949, *Phil. Mag.*, **40**, 481.

LETTERS TO THE EDITOR

Paramagnetic Resonance of Impurities in NaF

The paramagnetic resonance spectrum has been observed of a single crystal of NaF before and after irradiation; the results may be summarized as follows.

(1) Before irradiation. At 4°K a weak isotropic line about 50 gauss wide is observed, with $g=4\cdot28\pm0\cdot06$, arising from about 3×10^{14} centres.

(2) After irradiation with 75 megaroentgens of ^{60}Co gamma rays at 40°C. The crystals become a light pink colour and at 20°K a line with an isotropic g-value ($g=4\cdot344\pm0\cdot002$) appears with a well resolved structure. Analysis of this structure indicates that it arises from interaction of the magnetic electrons with nuclear magnetic dipoles of the fluorine ions. The intensity corresponds to about 2×10^{15} centres in a crystal of 0·05 cm^3 volume. Similar effects are observed after irradiation with 150 kv X-rays.

(3) After irradiation by 5 μA of 1 MeV electrons for 2 minutes. The same line is observed with enhanced intensity, and in addition there is a weak line with fluorine structure with an isotropic g-value of $1\cdot998\pm0\cdot002$. This line can be recorded at 90°K and arises from about 3×10^{14} centres.

(4) After irradiation with pile neutrons at 80°C (total dose, $1\cdot4\times10^{18}$ neutrons cm^{-2}). The pink coloration (associated with an intense absorption band in the region λ5100–5500 Å) is pronounced, but no resonance lines with a resolved fluorine structure are observed.

(5) The pink coloration and the irradiation-produced resonance lines are both annealed out by heating at 140°C for 48 hours, and could not be annealed out separately. Nevertheless, it seems unlikely that the two effects are directly associated since one appears without the other in the pile-irradiated sample; also, the number of magnetic centres is too small to give such an intense coloration.

(6) The broad line remains visible after annealing, and on re-irradiation the resolved structure at $g=4\cdot344$ reappears with the same intensity.

The high isotropic g-value indicates that the magnetic centre must be an impurity ion of one of the transition groups, and it must be situated in a position with cubic symmetry. The value is strikingly close to that of the Co^{2+} ion in MgO, which gives $g=4\cdot29\pm0\cdot01$, but with a large eight line hyperfine structure ($A=97\pm2\times10^{-4}$ cm^{-1}) (Hayes and Orton, unpublished). These results are consistent with the theory for Co^{2+} under octahedral symmetry, and since MgO has the NaF structure, they suggest that the magnetic centre in NaF is an isoelectronic ion (d^7) with no nuclear spin, on a cation site. The most plausible ion is Fe$^+$, whose ionic radius is estimated to be very close to that of Na$^+$; similar effects were observed with irradiated crystals to which iron in 0·01% molar concentration had been added in the melt.

Analysis of the structure shows that it must be due to hyperfine interaction with six fluorine nuclei located in a regular octahedron about the magnetic centre, confirming its location on a cation site. In the notation of Tinkham (1956), fluorine hyperfine components occur at fields displaced from the centre by

$$\delta H_z=\sum_N A_z^N I_z^N=\sum_N I_z^N\{A_s^N+A_p^N(3\cos^2\theta_{z,\sigma}-1)\}$$

where the summation is over six fluorine nuclei. Here it is assumed that the magnetic electrons are partly in pσ orbitals on the fluorines, and that the external field H_z is applied along the z-axis at an angle $\theta_{z,\sigma}$ with the axis of each σ-bond. A_s is determined by the s electron contact interaction and A_σ includes the dipole–dipole interaction and the coupling through pσ orbitals. It is expected that A_s and A_σ will be the same for all six fluorines. When H_z is along a {111} axis, the spectrum (figure 1 (*b*)) contains seven equidistant components with relative

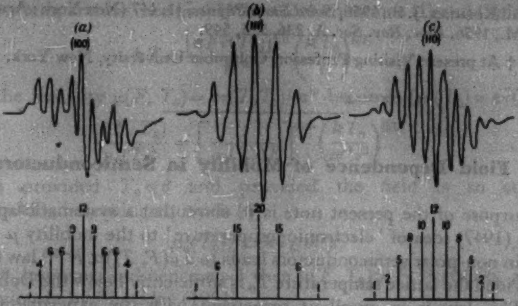


Figure 1. Fluorine hyperfine structure in chart recording of differential of absorption lines. The patterns for the {100} and {110} directions have been constructed using the approximate ratio $A_\sigma/A_s = 2/5$, which gives simple ratios for the line separations.

intensities 1 : 6 : 15 : 20 : 15 : 6 : 1, and line width 5 gauss. This is in agreement with expectation, since in this orientation $(3\cos^2\theta_{z,\sigma} - 1) = 0$ for all the fluorines, and only A_s contributes to the splitting; the separation of the lines gives $A_s = 14{\cdot}0 \pm 0{\cdot}6$ gauss $= (28{\cdot}3 \pm 1) \times 10^{-4}$ cm^{-1}. The spectra for the {100} and {110} directions (figures 1 (*a*), 1 (*c*)) are in good agreement with the calculated patterns, and we may take $A_\sigma = 5{\cdot}0 \pm 0{\cdot}5$ gauss $= (10 \pm 1) \times 10^{-4}$ cm^{-1}. These values are in the range found by Tinkham for Co^{2+} in ZnF_2.

Spectroscopic analysis showed iron to be present in concentration 2 parts per million, with chromium and nickel at half this concentration. Comparison with the magnetic resonance intensity indicates that in the irradiated sample substantially all the iron is present as Fe^+ ions, while before irradiation or after annealing less than 10% is in this form, if we ascribe the broad line at $g = 4{\cdot}28$ to such ions. The remainder may then be in the Fe^{2+} form, though no trace of a spectrum due to Fe^{2+} ions was found. However, the presence of ions of the wrong charge may account for the broadening of the Fe^+ line, and for the absence of an Fe^{2+} line, for which the broadening would be more severe. It may also account for the rather surprising absence of a well developed F-centre spectrum, whose formation is rather sensitive to such ions (Seitz and Koehler 1956).

The fluorine structure on the $g = 1{\cdot}998$ line indicates that this magnetic centre is also on a cation site, but the intensity is too small for a satisfactory analysis. It is possible that this line is due to Cr^+ ions; a detailed investigation of the resonance spectra of 3d ions in NaF and NaCl is now in progress.

We are indebted to Dr. D. A. Jones of the crystal growing section, Aberdeen University, for the crystals, and to the Atomic Energy Research Establishment, Harwell, for the radiation facilities.

Clarendon Laboratory,
Oxford.
21st March 1957.

B. Bleaney†
W. Hayes.

Seitz, F., and Koehler, J. S., 1956, *Solid State Physics*, II, 447 (New York : Academic Press).
Tinkham, M., 1956, *Proc. Roy. Soc.* A, **236**, 535, 549.

† At present Visiting Professor, Columbia University, New York.

Field Dependence of Mobility in Semiconductors

The purpose of the present note is to show that a systematic application of Fröhlich's (1947) idea of 'electronic temperature' to the mobility μ of electrons and holes in non-polar semiconductors leads to a $\mu(F, T_0) \propto F^{-0\cdot 8}$ law (F= electric field), provided the lattice temperature T_0 is sufficiently below the Debye temperature θ. This law is in excellent agreement with the experimental evidence obtained by R. Bray (private communication).

The idea is based on the assumption that inter-electronic collisions are so frequent that inter-electronic equilibrium is established faster than electron–lattice equilibrium. It has been shown (Fröhlich and Paranjape 1956) that this takes place at very much smaller electronic densities than superficial considerations might suggest. In an external field then the electrons will have a temperature $T(F) > T_0$, which increases with field strength.

A rough derivation will be given, ignoring numerical factors of order unity, and assuming a free electron distribution energy spectrum modified by an effective mass m^*. Let $\tau(T, T_0)$ be the average time of relaxation of electrons with energy $\boldsymbol{k}T$ assumed to be considerably higher than $\boldsymbol{k}T_0$. Thus since in a stationary state the rate of energy transfer from the field to the electrons must be equal to that from the electrons to the lattice vibrations

$$\frac{\mathbf{e}^2\tau(T, T_0)F^2}{m^*} = \frac{(\boldsymbol{k}Tm^*s^2)^{1/2}}{\tau(T, T_0)(1+2\bar{n})} . \qquad \ldots\ldots(1)$$

For, as explained by Fröhlich and Paranjape (1956), the average energy exchange per collision is $(\boldsymbol{k}Tm^*s^2)^{1/2}$ (s = velocity of sound, interaction with the acoustic vibrations only is considered); also, of $1+2\bar{n}$ collisions only $\bar{n}$ lead to absorption and $1+\bar{n}$ to emission of lattice phonons. Here

$$\bar{n} = [\exp\{(\boldsymbol{k}Tm^*s^2)^{1/2}/\boldsymbol{k}T_0\} - 1]^{-1} \qquad \ldots\ldots(2)$$

is the average number of phonons (per normal mode) which can interact with electrons of average energy $\boldsymbol{k}T$. The calculation of $\tau(T, T_0)$ follows the usual lines leading to

$$\frac{1}{\tau(T, T_0)} = \frac{1}{\bar{\tau}(T_0)} \frac{\boldsymbol{k}T(m^*s^2)^{1/2}}{(\boldsymbol{k}T_0)^{3/2}}(1+2\bar{n}) \qquad \ldots\ldots(3)$$

where $\bar{\tau}(T_0)$ is the low field relaxation time which is independent of F. Clearly as $T \rightarrow T_0$, $\tau(T,T_0) \rightarrow \bar{\tau}(T_0)$ since down to temperatures $T_0 \sim 1$°K

$$(\boldsymbol{k}T_0 m^* s^2)^{1/2} \ll \boldsymbol{k}T_0 \text{ so that } \bar{n} \gg 1.$$

As T increases with increasing field strength, $\bar{n}$ decreases. It follows from Fröhlich and Paranjape (1956) that $\boldsymbol{k}Tm^*s^2$ cannot be larger than $(\boldsymbol{k}\theta)^2$ because otherwise the crystal breaks down. Thus, if $T_0 \gg \theta$, then $\bar{n} \gg 1$ always holds. For $T_0 \ll \theta$, however, $\bar{n} \ll 1$ in fields sufficiently strong but below breakdown.

In the low temperature case ($\bar{n} \ll 1$) therefore equations (1) and (3) lead to

$$\boldsymbol{k}T = \frac{[\boldsymbol{e}F\bar{\tau}(T_0)]^{4/5}}{m^*} \left(\frac{\boldsymbol{k}T_0}{s}\right)^{6/5} \qquad \ldots\ldots(4)$$

and hence the mobility $\mu(F, T_0) = \boldsymbol{e}\tau(T,T_0)/m^*$ becomes $(\mu_0(T_0) = \boldsymbol{e}\bar{\tau}(T_0)/m^*)$

$$\frac{\mu(F, T_0)}{\mu_0(T_0)} = \left[\frac{s}{\mu_0(T_0)F}\right]^{4/5} \left(\frac{\boldsymbol{k}T_0}{m^*s^2}\right)^{3/10}. \qquad \ldots\ldots(5)$$

This holds provided $T_0 \ll \theta$ and provided the field is so strong that $(\boldsymbol{k}Tm^*s^2)^{1/2} > \boldsymbol{k}T_0$, i.e., using equation (4),

$$\mu_0(T_0)F > \boldsymbol{k}T_0/m^*s. \qquad \ldots\ldots(6)$$

When $T_0 \gg \theta$, however, and hence $\bar{n} \gg 1$ then $\mu(F, T_0) \propto F^{-1/2}$ results, as has been derived previously by Shockley (1951).

It should be mentioned that $\boldsymbol{k}T$ may be very much larger than $\boldsymbol{k}T_0$; its limiting value $[(\boldsymbol{k}\theta)^2/m^*s^2]$ at breakdown is of the order 1 ev. Hence the effective mass m^* occurring in (5) and (6) is only the one measured in normal weak field conditions, if the energy spectrum of electrons is of the free-electron type all through this energy region.

The author wishes to thank Professor H. Fröhlich for advice in all stages of the work.

Department of Theroetical Physics, B. V. Paranjape.
University of Liverpool.
28th March 1957.

Fröhlich, H., 1947, *Proc. Roy. Soc.* A, **188**, 521.
Fröhlich, H., and Paranjape, B. V., 1956, *Proc. Phys. Soc.* B, **69**, 21.
Shockley, W., 1951, *Bell. Syst. Tech. J.*, **30**, 990.

Note added in proof.—Dr. R. Stratton informs me that he has derived similar results.

REVIEWS OF BOOKS

Gaseous Nebulae, by L. H. ALLER (vol. 3 in The International Astrophysics Series). Pp. xvi+322. (London: Chapman and Hall, 1956.) 63*s*.

Some thirty years have elasped since the publication of Zanstra's quantitative theory of nebular luminosity and of Bowen's identification of forbidden transitions in the gaseous nebulae. The appearance of a monograph on these nebulae by one of the most active workers in the field is to be welcomed not only because of the clear account of recent advances which it contains, but also because of its comprehensive review of what is now known about the physics of rarefied gases in the presence of a dilute radiation field.

On the observational side the last thirty years have witnessed three notable advances. First, Bowen and Wyse, later joined by Minkowski and Aller, have extended the nebular spectrum to very faint lines by a happy use of Bowen's 'image-slicer', and have shown that there is a close quantitative similarity between the chemical composition of the nebulae and of the sun. Secondly, accurate line intensities and isophotes have been obtained for some of the brighter planetary nebulae. Finally, with a multiple-slit spectroscope O. C. Wilson has greatly extended the work of Campbell and Moore on internal motions in these nebulae and, among other interesting results, has confirmed Zanstra's suggestion that the nebulae are expanding.

On the theoretical side there has been no less activity. In a long series of papers Menzel and his colleagues have developed interesting methods for the interpretation of line intensities, and these have been admirably summarized in the fourth chapter of the monograph. It is useful to have these methods for the interpretation of line intensities. In the present state of knowledge of interstellar absorption, however, Zanstra's method of finding the ratio of line intensity, integrated over the nebula, to the intensity at the same wave length in the continuous spectrum of the exciting star seems more likely to yield exact knowledge of physical conditions in the nebulae. Following Bowen's original identification a great deal of theoretical work has been done on electric quadrupole radiation, and many astronomers will find the chapter in which Dr. Aller describes this work the most valuable in his book.

The monograph is written primarily for the astrophysicist who wishes to carry out research on the nebulae. The references are remarkably complete and the summaries of the work in these papers so clear that in most cases it is unnecessary to refer back to the originals. Equally, the book can be of value to the physicist who wishes to familiarize himself with the remarkable conditions prevailing in the nebulae, and to the astronomical student. For these readers the monograph might, however, have provided easier reading if the gaseous nebulae had been presented less as a subject in which our knowledge is complete, which incidentally is far from being the case, and more as one with knowledge growing in time. There are few branches of astrophysics where a single theory has been dominant so long as has Zanstra's, and few in which isolated pieces of work such as the Lick volume on nebulae in 1918, Hubble's investigation of the diffuse nebulae and Bowen's discovery of forbidden transitions have proved so dramatic and enduring landmarks of progress.

H. H. PLASKETT.

Optics and Sound, by C. A. PADGHAM. Pp. iv+434. (London: Pitman, 1955). 16*s*.

This book is one of the volumes of the Intermediate Science Series designed to meet the needs of students at the stage of the advanced level of the General Certificate of Education and of the University Intermediate course. The subjects are treated clearly and concisely and readers will find the work a useful and sufficient guide. At the end of most of the chapters a set of exercises is given which will be found helpful as a test and as a revision. A good feature of the book is that the dates of the principal contributions to Optics and Sound, mentioned in the text, are given after their names.

A disappointing feature of young physicists is their ignorance and disregard of the simplest historical facts concerning the pioneers of their subject. It is a good beginning at this early stage to record at least the period of their lives.

At this level Optics is by custom and, to some extent, of necessity chiefly concerned with the geometrical aspect. The author has done well in choosing the Cartesian system for his sign convention. He has stated it clearly and illustrated it in examples. It may be hoped that teachers and other writers will follow this example. The confusion over sign conventions at this level and even later has contributed to a distaste for this subject at a more advanced stage. There is only one serious rival to it: the convention attributed to Abbe, and it is a very simple matter to pass from one notation to the other. This convention has the drawback from the point of view of elementary teaching that in the simple case of the concave mirror certain magnitudes seem to be unnecessarily associated with negative signs.

In this book the reader is introduced to the sign convention first through refraction at spherical surfaces. Reflection at spherical surfaces is treated later. There appears to be no strong reason against the adoption of this change in the customary sequence and there is much to be said for it, if it contributes to the adoption of this sign convention.

The section on Optics ends with a short chapter on Photometry and one on the theory of certain natural phenomena, polarization and interference. These are adequate for students at this stage. However, the author misses the opportunity of contrasting the application of the principle of superposition in Optics with that to which he refers later in the section on Sound. In the latter case he might have stressed the fact that it applies to vibrations of small amplitude. There is a statement in the paragraph on polarized light which might be omitted in the future. The author writes: " It is an interesting fact that a modification of Newton's theory, the modern quantum theory, must be used today, together with the wave theory, to explain all known phenomena of light ". There are features common to Newton's corpuscular theory and to modern quantum theory but it is misleading to describe the modern quantum theory as a modification of Newton's theory.

The short section on Sound gives the main features and principles of the subject suitable for the student at this stage of his career. The subject tends to be brushed aside at the Intermediate level but here the author keeps the subject alive without omitting the fundamentals at the expense of more recent and popular applications.

This is a book to be recommended for its conciseness and clarity in the field for which it is written. H. T. FLINT.

Progress in Semiconductors, Volume 1, by A. F. Gibson, P. Aigrain and R. E. Burgess. Pp. vii+220. (London: Heywood, 1956.) 50s.

One must agree with the editor that, with the vast amount of research at present being carried out in the field of semiconductors, there is need for a book which contains survey articles of the fields of investigation, and that such a volume should be published annually.

This volume contains seven articles, the first of which deals with the technically important semiconductor, silicon. The article introduces all the important aspects of research on this material from its preparation in a state of very high purity, through results of fundamental experiments carried out on the raw material, to the properties of devices made from it. Of necessity the author does not delve deeply, but the bibliography is extensive and lists most of the important papers on these subjects, though several interesting references are tantalisingly labelled " private communication ".

The second article provides a readily comprehensible introduction to the application of 'filament' techniques. Although the brevity of the present treatment precludes mathematical analyses of the experimental situations, the well-known techniques are discussed, together with the more recent work on such phenomena as carrier extraction, etc. For those requiring detailed information on a particular topic, a bibliography is appended.

A very good review of the theoretical work on the Seebeck effect is given in Article 3, though a reader not already well acquainted with the subject might find the paper difficult to read since several not very obvious assumptions are made. For the non-specialist a fuller definition of the properties under discussion, and a list of symbols, would have improved the understanding of the article.

In the article on phosphors, the author first outlines the general characteristics of sulphide phosphors and explains the possible modes by which a phosphor system may be excited. Subsequent sections deal in some detail with the phenomena of photo-conduction, dielectric changes, semiconducting properties of phosphors, and electroluminescence, and provide a fairly thorough review of the work published on these materials in the last eight years.

For the non-specialist, the article on the Design of Transistors achieves its purpose in giving a descriptive account of the factors limiting their performance at high frequencies, together with a survey of those devices being developed to overcome the limitations. No mention is made of the conflicting design problem arising because a decrease in ρ_b involves an increase in C_c, and an inconsistency exists between equation 3 and equation 4, due to a difference in treatment. The article was written before the recent developments using the powerful technique of impurity diffusion were publicized.

The 5th article, for the specialist, gives a good account of the theory of the photo-magneto-electric method for measuring lifetimes and surface recombination velocities in semiconductors. Assuming simple recombination of added carriers, the authors derive expressions for photo-magneto-electric voltages, and the phase shift between the photo-magneto-electric and photoconductive voltages. However, no mention is made of the possible effect of the temporary trapping of carriers on the a.c. analysis; such an approach would have been valuable in view of the possible application of this method to silicon and other semiconductors.

The last article is probably the most disappointing one since the title leads the reader to expect a more general account of the characteristics of surface states. In fact, the author presents his work on the surface states of tellurium, and in dismissing in one line the experimental work on germanium and silicon fails to introduce interesting features—fast and slow surface states—which have arisen from these studies. In a review article, a better balance should have been maintained between the author's work and that of workers on other materials.

On the whole, the book fulfils its objects as given by the Editor in the preface. There are omissions— semiconductor work in Germany and work on the III-V compounds—but the Editors will have these subjects in mind for future volumes.

I. WILLIAMS.

Modern Physics: A Textbook for Engineers, by R. L. SPROULL. Pp. xii+491. (New York: John Wiley; London: Chapman and Hall, 1956). 62*s*.

Perhaps the greatest difficulty encountered in planning or revising an undergraduate course in science or engineering is to decide how much concerning recent developments can be included without unduly overburdening the student. This is especially the case in electrical engineering where a good deal of modern physics is demanded if the undergraduate is to be equipped with a knowledge of the underlying principles involved in such post-war innovations into industry as modern magnetic alloys, nuclear engineering and semiconductor devices. This book by Professor Sproull is, therefore, welcome in that it is the outcome of experience gained in "a one-semester course in atomic, solid state and nuclear physics" for engineering students at Cornell University.

In the preface the author states: "The heart of modern physics is quantum mechanics, and the central part of this book, is devoted to treating it". It cannot be said, however, that the introduction given to quantum mechanics is successful. The Schrödinger wave equation is stated without any kind of derivation after the assertion that "this equation replaces $\mathbf{F}=m\mathbf{a}$ for the motion of particles on an atomic scale of sizes", and a few general remarks, without mathematics, on wave equations and the important probability concept. Again, the four quantum numbers n, $\mathbf{l}$, $\mathbf{m}_l$ and $\mathbf{m}_s$ are introduced with inadequate justification, despite their importance in the theory of atomic spectra and in the enunciation of the exclusion principle. The good student will here be unsatisfied and the mediocre student will be merely mystified. Similar inadequacies mar the accounts of the theory of the solid state and the chapter on semiconductors. One wonders whether the author has been wise in rejecting too readily much of the older classical quantum approach in favour of the wave mechanical treatment. Admittedly, the older quantum theory has to be revised considerably in the light of wave mechanics, but it does provide the student with an essential background against which the newer concepts can be described with success.

It is, however, too easy to criticize textbook accounts of modern physics, especially when they are intended for the engineering student faced with studying physics as an ancillary subject. Though parts of this book would prove difficult to understand without considerable amplification in lectures, yet it is well written, excellently illustrated, contains many interesting examples, and provides a good coverage of those aspects of the theory and practice of modern physics which the well-equipped electrical engineer should know, even though he may find it preferable to leave much of the study to postgraduate courses.

J. YARWOOD.

Molecular Beams, by N. F. RAMSEY. Pp. xii + 466. (Oxford: Clarendon Press; London: Cumberlege, 1956). 75*s*.

Much has been said about the usefulness and elegance of the method of molecular beams. Several Nobel prizes have been awarded to workers in this important field, and Professor Ramsey himself has contributed a great deal both to the experimental and to the theoretical development of the subject.

This is the first full-length book to be written about molecular beams since Fraser's *Molecular Rays* appeared 25 years ago. It was not until 1938 that Rabi and his associates introduced the method of radio-frequency resonance, which paved the way for some of the most precise and fruitful physical measurements ever made. Professor Ramsey's book is devoted mainly to the consequences of this development. Nevertheless, it covers in a detailed and consistent manner the entire subject, including the material of Fraser's book.

The substance of the latter is discussed in the early chapters on gas kinetics and non-resonance measurements of magnetic moments. The theory of the interaction of a nucleus with atomic and molecular fields is clearly set down, and then follows a discussion of the magnetic resonance method and its application to molecular, atomic and neutron beams. Tables of nuclear magnetic and electric moments, hyperfine structure separations and molecular constants (up to 15th September 1955, with notes added in proof) are included in the appropriate chapters, together with theoretical interpretations of these data. Separate chapters are devoted to electric deflection experiments and to the Lamb-Retherford experiments on atomic fine structure. Finally, there are two chapters on design principles and techniques, which serve as a laboratory guide to research workers, though, as the author points out, these cannot, and are not intended to, replace first-hand experience in an established molecular-beam laboratory. Not the least useful is a section containing over 700 references.

There are a very few misprints and occasionally a wild guess at authors' initials, but the many drawings are admirably presented and harmonize well with the text.

This is essential reading for graduate students specializing in the field of molecular beams, and will be invaluable, if only as a reference book, for more experienced workers; and because it is bound to become a standard work, it cannot be ignored by a much larger class of physicists. G. K. WOODGATE.

Journal of Fluid Mechanics, edited by G. K. BATCHELOR (bi-monthly). (London: Taylor and Francis.) £5 10*s*. per vol.

Although this is an era of specialization it has also seen the gradual merging of the various sciences, and fluid mechanics, the subject under review, is where the fields of aerodynamics, hydromechanics, meteorology and physics have overlapped. The arrival of this new journal is therefore particularly opportune, and an editorial board which includes the names of Batchelor, Lighthill, Benjamin and W. C. Griffith, should ensure the maintenance of a high standard. In the first issue there are eight papers of average length—about fifteen pages—and on subjects ranging from collisions of drops in clouds to supersonic flow and boundary layer phenomena.

If a single paper may be taken for comment, that of Blackman on vibrational relaxation is indicative of where the results of physical experimental investigations are having their impact on the theoretical work in this field. It is to be hoped by the reviewers that future issues will give us more papers of this type than are in the first issue which is biased on the theoretical side. A commendable impression of this issue is the clarity of presentation which augurs well for the future of the journal.

H. J. P.
R. W. B. S.

Irradiation Colours and Luminescence, by K. PRZIBRAM. Translated and revised in collaboration with the author by J. E. Caffyn. Pp. xiv + 332. (London: Pergamon Press, (Pergamon Science Series. Physics, Vol. 1), 1956). 63*s*.

It is surely rare if not unique for a scientist of 77 to be still active in a field in which he has laboured with notable success for some 40 years and to be vigorous enough to collaborate in a translation, revision and expansion of his book published in German in 1953, and it is fortunate that to his own knowledge of the subject and of English there has been added the help of an English worker who has had considerable experience in the same field.

The result is a readable and (as far as the present reviewer is aware) the most complete and up to date account of the natural and artificial coloration of inorganic crystals and of the luminescent and some electrical properties associated with the coloration. To the physicist and chemist the main importance of the work described and discussed would seem to be the impressive attempt to establish that the coloration of many minerals and artificially produced crystals (a list of some two dozen is given) is often due to radioactive influences, either directly, or indirectly through valency changes produced in impurities, is structure sensitive and impurity sensitive and often shows a morphology of significance for the formation, history and location of minerals. Clearly the geologist will find here many matters of interest and importance.

The evidence that the changes of physical properties which accompany the coloration are closely related to the nature of the colour centres and to the physical processes involved in their formation and transformation is briefly marshalled. There is some confusion in the discussion of F centres and a surprising omission of the theoretical work of Simpson. A very full treatment is given of the coloration of rocksalt and of fluorite with shorter accounts of other halides, of oxides, sulphides, nitrates, phosphates and silicates and a brief but important chapter on pleochroic haloes. Considerable service has been done by the revival of many old half-forgotten observations which are now seen to take on a new significance in the light of modern ideas about the solid state. The facts presented and the discussions both provide many suggestions for future work. It has recently been suggested that one of the most important outstanding problems in physical science is the discrepancy between the theoretical and actual strength of solids. A reading of this book cannot fail to suggest that the effect of radiations may well play a part in the elucidation of this problem. The 49 pages of references will be found immensely useful. Two complaints must be recorded. There are many faulty references in the text to the figures illustrating it and some lack of clarity in the meaning of some of these figures.

J. EWLES.

Principles of Guided Missiles Design. Vol. 1 : *Guidance*, by A. S. LOCKE. Pp. ix + 748. (New York: Van Nostrand ; London: Macmillan, 1955). 90*s*.

This book is the first of a four-volume series intended to cover all aspects of guided missile technology to a standard appropriate to graduate students and professional engineers.

A treatise of this nature must be severely handicapped by the necessity to avoid revealing classified information and a book of 713 pages on the unclassified aspects of missile guidance must be expected to present in considerable detail all the basic technologies that are used in, but are by no means particular to, guided missile systems.

The author devotes considerable sections of the book to terrestrial and celestial reference systems, to the basic principles of transmission of radio wave propagation, to the theory of mathematical transforms and to servo system theory before dealing with the characteristics of missiles and missile systems. Some of these chapters terminate in a useful bibliography ; an unfortunate exception, however, is the chapter on detection and information gathering. Thirty-two footnote references are supplied within the chapter but the consolidated bibliography is omitted.

The sequence of chapters is somewhat surprising. Chapter 8 considers some hypothetical operational requirements and discusses the technical means of fulfilling them. The currently available methods of missile guidance are not discussed and defined, however, until Chapter 16 and the pages following page 539; an exceptionally late stage to define the material that is the primary purpose of the book.

The author emphasizes throughout the importance of treating the complete weapon system as an integrated whole and stresses the importance of bandwidth considerations in assessing system performance.

There are a few instances of statements that may confuse the uninitiated. For example, the suggestion that a wide-angle primary source is the major cause of side lobes in a radar antenna system. There is also confusion between circular and hyperbolic navigation systems. Nevertheless, the book can be welcomed as succeeding in a difficult task and as providing a useful and informative introduction to the young scientific or engineering graduate who is starting a career in the guided missiles field. C. HOLT SMITH.

Flow of Gases through Porous Media, by P. C. CARMAN. Pp. ix + 182. (London: Butterworths Scientific Publications, 1956). 30*s*.

Dr. Carman has provided a valuable addition to the literature of the subject by writing his book on the flow of gases through porous media. To collect together in this accessible form the knowledge of a subject which has developed rapidly in recent years is a considerable achievement. Several hundreds of references are given and most of these relate to the last fifteen years, as Dr. Carman has pointed out. The work will be particularly useful to those who are engaged in the application of science to practical problems. It will also be of value to all who are interested in gas kinetics.

Dr. Carman deals first with viscous flow through beds of discrete particles of various forms, followed by a short account of consolidated media. Free-molecular or Knudsen flow, where the mean free path is greater than the molecular diameter, is dealt with fully and carefully. An account is given of the permeability method of surface area measurement, a subject to which the author has himself made a number of contributions.

There is a useful chapter on the separation of gas mixtures by flow and diffusion in porous systems which is now a matter of great practical importance.

The text maintains a nice balance between theory, experiment and practical applications. The book will be of interest to many scientific workers and, as the subject will continue to develop, it will no doubt be needed in further editions. When this occurs I suggest the addition of a simple introductory chapter on the lines of the present preface but considerably amplified. This will enable those who are not already familiar with the subject to pick up the essential threads more easily before meeting with numerous equations. T. L. IBBS.

Automatic Digital Computers, by M. V. WILKES. Pp. x+305. (London: Methuen, 1956.) 42*s*.

This book contains a great deal of information about automatic digital computers, their historical development, design, and use. The book will be of interest primarily to persons who wish to understand how computers work, but it will also be read by those interested in their applications. In the early post-war period the designers and builders of machines were very often the people most interested in their use. Nowadays, however, the distinction between builder and user, that is, between engineer and mathematician, has become so marked that the majority of persons trained in the use of a machine have little appreciation of the details of its design beyond what is necessary to programme for it. Nevertheless, it is essential for the successful design of new machines to combine the know-how of the engineer and the mathematician. This particular task is known as *logical design*, and it is the author's experience in this field that makes the book an important contribution to the literature.

The chapter headings are:—(1) The development of automatic digital computers (38 pages); (2) The principles of logical design (47 pages); (3) The principles of programme construction (47 pages); (4) Relay computers (19 pages); (5) Storage (58 pages); (6) Electronic switching and computing circuits (47 pages); (7) The design and operation of digital computers (32 pages). There is a short appendix on 'Machinery and Intelligence' and a very useful bibliography.

The opening chapter gives an account of the work of Charles Babbage, and descriptions of the Automatic Sequence Controlled Calculator, the ENIAC and the EDVAC. An account of the early relay computers is given later on in the book. In the next two chapters the principles of logical design and programme construction are developed, with particular reference to the EDSAC, but at the same time making adequate reference to other machines. The various input and output systems employed with computers are also described. The chapter on storage covers all known practical systems, ultrasonic, electrostatic, and magnetic, including the most recently developed static magnetic matrix storage. Symbolic switching algebra is introduced for the discussion of binary and decimal computing circuits: however, very few detailed circuits are given, and no mention is made of the use of transistors.

The final chapter deals with the problems of the engineer charged with the overall responsibility for the design and construction of a computer. Thus it discusses the relative importance of such desirable features as speed of operation, reliability, ease of maintenance, ease of programming, etc. The problems of running a computing centre, such as staffing and organization, are also discussed, and the chapter concludes with an account of the use of digital computers for business purposes and the control of industrial processes. This chapter alone makes the book valuable, since so little has been written on these particular topics. R. A. BROOKER.

Science and Information Theory, by L. BRILLOUIN. Pp. xvii + 320. (New York: Academic Press; London: Academic Books, 1956.) $6.80.

In recent years Information Theory must have been a source of uneasiness for many physicists, as one of the subjects ' on which one hears so much and which one ought to read up '. Here at last is a book which one can wholeheartedly recommend to all physicists and scientists of an advanced level. It is written by an outstanding physicist, who has left his mark on such widely different fields as Brownian motion, light propagation, quantum theory, the physics of solids, ultrasonics, thermodynamics, waveguides, short wave electronics and computers. Nothing less than this unique combination of specialist experiences was required to write this remarkable work.

The central theme of the book is the relation of the process of collecting information on physical systems to the second principle of thermodynamics. When Claude Shannon, in 1948, wrote down his measure of information per communication symbol as the negative average of the logarithms of the symbol probabilities, and named it ' source entropy, H ' (because of its formal identity with Boltzmann's H-function which is the *negative* of entropy), he can hardly have foreseen the confusion which it was to create in the heads of electrical engineers, the chief users of communication theory, who could not be expected to know much of thermodynamics, or else he would probably have chosen another name for it. In the end it turned out to be a good term, apart from the regrettable negative sign. Unknown to Shannon, L. Szilard had found, twenty years earlier, the same mathematical expression for the minimum cost of entropy increase which one has to pay for collecting information on a physical system. More than half of Brillouin's book is devoted to the elucidation of the by no means quite simple connection between information and entropy, which he has widely generalized and brilliantly illustrated. It can now be said that the principle of indeterminacy and the ' negentropy ' principle are the two fundamental limitations of all physical measurements, but while the importance of the first is long recognized, the second has not so far received all the attention which it deserves. I regret only that (apart from a few brief remarks on p. 233) the author has not emphasized that the exorcism of the Maxwell Demon ' incidentally ' proves also the inconsistency of classical physics with thermodynamics, that is to say that the second principle, like the third, derives its validity from quantum physics. This is a contention first put forward by Max Born, and not yet universally accepted.

Though the book will be most enjoyed by advanced physicists, it can be equally recommended to all scientists who had no previous contact with information theory, and to electrical engineers who had no intimate knowledge of statistical physics or with thermodynamics. The main results of communication theory are all derived from first principles with just that level of mathematics which theoretical physicists have long accepted as their standard, that is to say with careful, even painstaking, discussions of consequences, but without digging down into the foundations as deep as set theory.

The wide field covered by the book can be judged by its main chapters: Definition of Information. Redundancy in the English Language. Coding, Channel Capacity, Error Correcting Codes. Analysis of Signals. Summary of Thermodynamics. Brownian Motion. Thermal Noise in Circuits. The Negentropy Principle. Maxwell's Demon. Physical Limits of Observation. Writing, Printing, Reading. Computing. (20 chapters in all.)

It is, as yet, hard to say what influence information theory will have on the future development of physics. Of all scientists the physicists are most reluctant to adopt statistical methods. The structural simplicity of their subject enables them to consider parameters one by one, and to reject conclusions obtained by elaborate statistical methods from complex material which does not show a clear result on inspection. But information theory, though it is in essence a statistical science, does not suggest giving up this noble prejudice. It suggests merely that in many cases a set of parameters can be more accurately obtained from complex experiments than from one-by-one tests. Moreover, the theory gives upper bounds for the legitimate number of questions which a given experiment can be expected to answer. Finally, information theory may have a heuristic value in modern theoretical physics, by forcing the investigator to have both feet on the ground.

The reviewer has two personal reasons for warmly welcoming this book. One is, that the author has embodied in it some of his privately circulated considerations on Maxwell's Demon, thus relieving him of the trouble of publishing. The other is, that this is the book which the reviewer has been pressed to write for several years, but he could never have done it as well as Professor Brillouin.

D. GABOR.

CONTENTS OF SECTION A

PAGE

Prof. J. M. Cassels, Dr. G. Fidecaro, Mr. A. M. Wetherell and Dr. J. R. Wormald. The Capture of Stopped Negative Pions by Hydrogen . . . 405

Dr. J. H. Carver, Dr. R. D. Edge and Mr. K. H. Lokan. Direct Excitation in the Photodisintegration of Tantalum above the Giant Resonance 415

Mr. J. L. Lloyd, Dr. E. Rössle and Dr. A. W. Wolfendale. The Scattering of High Energy μ-Mesons in Lead 421

Mr. T. H. R. Skyrme. Nuclear Moments of Inertia 433

Mr. R. Berenbaum and Mr. J. H. Matthews. Gamma Radiation following Proton Capture in ^{35}Cl, ^{37}Cl and ^{39}K : II—Angular Distributions and Correlations . 445

Dr. E. W. Kellermann, Dr. T. Shaw and Dr. N. Dickinson. On the Structure of Extensive Cosmic Ray Air Showers : Core Detection and Core Structure . 452

Mr. R. McCarroll. The Excitation of the Discrete Levels of Atomic Hydrogen by Fast Electrons 460

Dr. H. R. Moore. The L-Series Emission Spectrum of Krypton 466

Research Notes :

Mr. N. Lynn and Dr. B. L. Moiseiwitsch. The Mobility of He$^+$ Ions in Helium 474

Prof. N. Feather, Dr. N. Miller and Mr. S. W. Peat. Search for a Rare α-Emission in ^{228}Ra (MsTh 1) 478

Dr. J. Berkowitz. On a New Phenomenon in Activated Nitrogen . . . 480

Contents of Section B 484

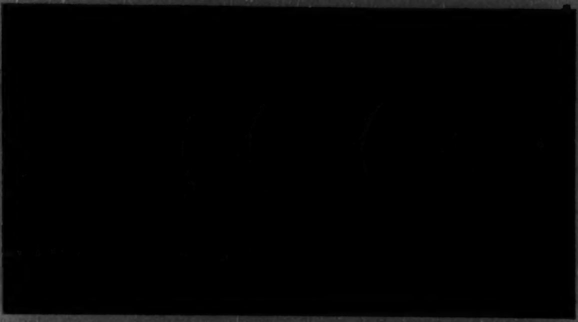


Figure 6. Micrograph showing slip lines on a crystal with a coating of gold 1500 Å thick, cadmium 1, $\chi_0=31°$, $\lambda_0=32°$, extension 9%, magnification ×27.

Figure 7. Micrograph showing the start of rupture of the gold film along the slip planes where the slip direction is parallel with the tangent of the specimen surface. The direction of view lies in the slip planes. Cadmium 1, $\chi_0=14{\cdot}5°$, $\lambda_0=15°$, thickness of gold film 12 000 Å, magnification ×23.

Figure 8. Micrograph showing rupture of the gold film after long extension, in a crystal of cadmium 3, with $\chi_0=51°$, $\lambda_0=51°$. The direction of view is nearly normal to the slip planes. Thickness of film 24 000 Å, magnification ×17.

The Magnetic Refrigerator for Temperatures below 1°K†

By J. G. DAUNT
Ohio State University

MS. received 21st *March* 1957

Abstract. The magnetic refrigerator is a device for maintaining isothermally a reservoir and/or experimental apparatus at temperatures below 1°K. Starting from temperatures maintained by a liquid helium bath the device produces cold by the magnetic cooling method, the cooling being performed in a cyclic manner. By cyclic operation temperatures below 1°K, as low as approximately 0·2°K, can be maintained continuously. In this paper some design problems are considered in detail and a comparison is made between the theoretically computed performance and the experimental observations on the first few magnetic refrigerators which have so far been constructed. The 'engine' of the magnetic refrigerator basically requires, besides a paramagnetic working substance for the magnetic cooling process, thermal valves which may allow or block the flow of heat between the working substance, the helium bath and the low temperature reservoir. In the work reported in this paper the material for the working substance has been ferric alum and the thermal valves have been constructed from superconducting lead wires. A discussion is given of the requirements for an ideal working substance and experimental results are reported on some unusual paramagnetic materials which have been investigated for use both as working substances and for reservoir materials. Amongst these ferric acetylacetonate and synthetic ruby have been found to show useful properties. Finally, some of the possible uses of the magnetic refrigerator are considered, as for example in calorimetry below 1°K, for the establishment of an isothermal reservoir for the initiation of single-stage magnetic cooling, for the production of liquid helium baths below 1°K and for the possible subsequent extension of the magnetic refrigerator to a two-stage device for the production of isothermal temperatures well below 0·1°K.

§ 1. Introduction

The magnetic method for the attainment of temperatures below 1°K, devised independently by Giauque (1927) and Debye (1926), has been in use in many laboratories since its first successful operation in 1933 (Giauque and Macdougall 1933, de Haas, Wiersma and Kramers 1933, 1934). This method makes use of the magneto-caloric effect in paramagnetic substances which, when adiabatically magnetized, are raised in temperature and conversely when adiabatically demagnetized suffer a temperature drop. Technically the process involves mounting a suitable paramagnetic substance in an evacuable chamber, which is kept in liquid helium at low temperature, and first magnetizing it in an intense magnetic field. This magnetization is in general made isothermal† by including exchange gas in the chamber, so allowing the heat of magnetization to be conveyed at constant temperature to the liquid helium bath. To ensure subsequent adiabatic demagnetization of the paramagnetic substance, the exchange gas is pumped out of the chamber and the magnetic field reduced, generally to zero. In the resulting cooling, temperatures as low as a few milli-degrees absolute can be reached within the paramagnetic substance, the final temperature being a function of the initial magnetic field and temperature and of the physical characteristics of the paramagnetic (for recent reviews of the magnetic cooling method see, for example, de Klerk and Steenland 1955, Ambler and Hudson 1955, Garrett 1954). Subsequently, due to unwanted heat influx from the warmer surroundings, the paramagnetic substance continuously warms up, until in general it finally reaches the temperature of the helium bath.

---

† Containing material included in the lecture delivered at the presentation of the Duddell Medal on 20th September 1956.

The maintenance of temperatures below 1°K continuously and isothermally can be effected by performing a cyclic magnetic cooling process which regularly extracts from a thermal reservoir heat sufficient to counterbalance the heat gained or developed in the reservoir. Such a cyclic process was devised by Daunt and Heer in 1949 (1949 a, b), at which time estimates of the performance of such a 'magnetic refrigerator' were given. The main requirement for this cyclic process lay in the use of thermal valves which could be opened or closed to the flow of heat at all temperatures of operation at and below 1°K. Successfully operating thermal valves were assured by previous observations (Heer and Daunt 1949, Heer 1949) on the heat conductivity of pure superconducting Sn and Ta at temperatures well below 1°K. It was found that at temperatures below 1°K, which for Sn and Ta were well below their transition temperatures, the ratio K_n/K_s of the thermal conductivity in the normal state to that in the super state could become very large (values of order 100:1 were observed). If therefore a suitable superconducting metal is changed from the super state to the normal state, as can be effected by the application of a small external magnetic field, the metal changes from being a poor thermal conductor to a good thermal conductor. This formed the basis for a magnetic thermal valve.‡ The basic mode of operation of a cyclic magnetic cooling engine, using such thermal valves, is described in § 2 below. It is noted that its successful operation was first reported by Heer, Barnes and Daunt (1953, see also Daunt 1953) in 1953 and a detailed description given by them in 1954. More recently a commercial model of the machine has been constructed by A. D. Little Inc. (Daunt, Heer, McMahon, Reitzel and Simon 1955).

---

† The presence or absence of exchange gas is not the only method whereby thermal contact between the paramagnetic substance and the helium bath may be made or broken. Thermal valves, performing the same duty, have been devised and used in the magnetic cooling method which depend for example on: (1) the difference in thermal conductivity in superconductors in the super and normal states, as first employed by Heer and Daunt (1949), (ii) the mechanical change in the dimensions of a channel of liquid helium, as developed by Ashmead (see Chase and Herlin 1955, Osborne 1956), (iii) the make and break of a metal mechanical contact, as used by Collins and Zimmerman (1953), and (iv) the thermally rectifying properties of columns of solutions of ^{3}He in ^{4}He, as first employed by Hwang, Fulton and Fairbank (1956). A proposal has been made by Mendelssohn and Rosenberg (1953) to use the magneto-resistive effect in cadmium single crystals as the basis for a thermal valve.

‡ At about the same time as this work was done, suggestions were put forward independently by Gorter (1948) and by Mendelssohn and Olsen (1950) for thermal valves using this principle of operation.

§ 2. MODE OF OPERATION

Figure 1 shows diagrammatically the general arrangement of the magnetic refrigerator. It consists essentially of a paramagnetic working substance A, a thermal reservoir R which is maintained continuously at the desired temperature below 1°K and to which experimental equipment can be attached, two superconducting thermal valves made of Pb ribbon, V_B being that between the working substance and the bath and V_R being that between the working substance and the reservoir. This system of 'engine' and reservoir is mounted in an evacuable chamber C which is immersed in a helium bath maintained at about 1°K. Surrounding the bath Dewars are the control magnets, all iron-clad solenoids with axes parallel to the Dewar axis. The centre magnet is for magnetization of the working substance in fields of from 7 to 9 kilooersteds and the two end solenoids which give fields of about 900 oersteds activate the 'opening' and 'closing' of the thermal valves. Detailed constructional details, including those for the magnets, have already been given by Heer, Barnes and Daunt (1954).

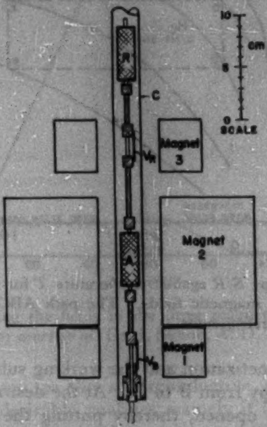


Figure 1. Diagram of general arrangement of magnetic refrigerator. A is the working substance, R the reservoir, V_B and V_R the thermal valves and C the vacuum jacket which is immersed in liquid helium.

The operation of the cycle follows as closely as possible that of a Carnot cycle and it can be followed by appeal to the entropy diagram of figure 2. The isothermal magnetization corresponds to the path A → B on the diagram during which the main magnetic field is raised from H_A to H_B. During this process, which takes about 40 seconds, the valve V_B is kept 'open' to the flow of heat from working substance to bath by maintaining its Pb element in the normal state by application of its magnetic field of 900 oersteds. Also during this process

V_R is 'closed'. At B, V_B is closed by reducing the magnetic field around it to a small value in a time of approximately 5 seconds, and then the main field is reduced to the value H_C. Since both valves V_B and V_R are closed this process is one of adiabatic demagnetization and the working substance therefore follows the line of constant entropy from B to C. At the desired low temperature at C the reservoir valve V_R is opened, thereby putting the working substance and reservoir in thermal contact and *at the same time* the main field is further continuously reduced from H_C to H_D (H_D is a few hundred oersteds in value). By suitable choice of the rate of reduction of this field from H_C to H_D the path C → D can be made quasi-isothermal, the cold generated by the demagnetization of the working substance counterbalancing the heat flowing into it from the reservoir. When the field has reached the near zero value H_D, the valve V_R is again closed by reducing the magnetic field around it to zero, thus isolating the reservoir once again. Then the main field is raised from H_D to H_A. This is an adiabatic process which therefore raises the temperature of the working substance once again to approximately that of the bath, the working substance following the isentropic path D → A. From here the cycle is repeated. The cycle therefore approaches a Carnot cycle consisting of two approximately isothermal and two approximately isentropic paths. The cycle moreover is controlled by the application and removal in correct sequence of three magnetic fields only, one for the generation of the magnetocaloric effect in the working substance and two to operate the thermal valves. The complete cycle takes approximately two minutes and a diagram showing the timing of the field switchings is given in figure 3.

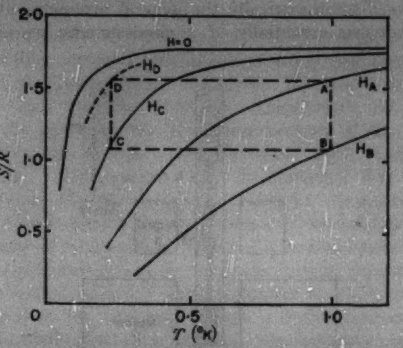


Figure 2. Plot of the entropy S/R against temperature T for the material of the working substance for various magnetic fields. The path ABCD represents the idealized cycle of the magnetic refrigerator.

§ 3. Choice of Thermal Valve Sizes

Detailed design and constructional details of the magnetic refrigerator have already been given by Heer, Barnes and Daunt (1954) and will not be repeated here. One problem which remains of some interest, however, relates to the choice of size for the thermal valves and the effect of such choice on the performance of the machine. It was found that the valve sizes required for successful operation were relatively critical and in consequence a further elaboration of the method of their calculation follows.

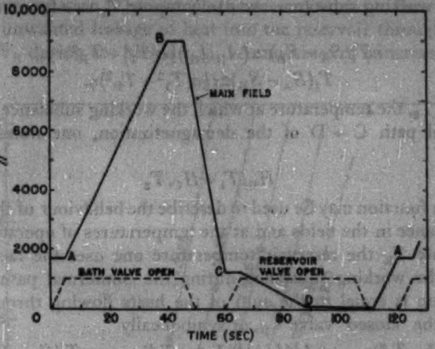


Figure 3. Diagram showing the fields in the three magnets as a function of time. (*Reproduced by courtesy of* Dr. I. Simon, A. D. Little Inc.)

Lead has been chosen as the superconductive material for the thermal valves since it has two desirable qualities, namely it is readily available in high physical and chemical purity and it has a high transition temperature. Both of these properties favour the highest values of K_n/K_s at any given temperature. The thermal conductivity of Pb in both the normal and superconductive state has been measured by a number of workers (de Haas and Rademakers 1940, Rademakers 1949, Nicol and Daunt 1952, Olsen and Renton 1952, Mendelssohn and Rosenberg 1952, Heer, Barnes and Daunt 1954). Although there is some scatter in the results of observations at different institutions, we have assumed (Heer, Barnes and Daunt 1954) that, over a limited temperature range from about 0·2°K to 1·1°K, K_n and K_s for Pb are given by

$$K_n = aT \simeq 4T \text{ w cm}^{-1} \text{ deg}^{-1}$$
$$K_s = bT^n \simeq 8 \times 10^{-2} T^3 \text{ w cm}^{-1} \text{ deg}^{-1}. \qquad \ldots\ldots(1)$$

This range of temperature of applicability of the above evaluation of K_n and K_s corresponds approximately to that covered by the thermal valves in the magnetic refrigerator. With these values of a and b, the performance of the machine for various thermal valve sizes can be computed in terms of the parameters n the number of moles of working substance in the engine and τ_{AB} and τ_{CD} the times for the isothermal magnetization A → B and isothermal demagnetization C → D (measured in seconds). The valve sizes are expressed by A_B/L_B and A_R/L_R in centimetres, where A is the cross-sectional area and L the length of the Pb element in each valve V_B and V_R. Since the calculations involve numerical computation, exact values for H_A, H_B, H_C and T_B, the helium bath temperature, must be specified for each computation.

The first step is the computation of the temperature T_1 of isothermal magnetization for the path A → B (see figure 2) in terms of a parameter $\eta = (A_B/L_B)\tau_{AB}/n$. During magnetization the heat developed flows largely through V_B down a temperature drop of $T_1 - T_B$ and a little through V_R from T_1 to T_R. Knowing a and b (see equation (1)) and the entropy diagram for the paramagnetic working substance, we can compute T_1 as a function of η from the relation†

$$nT_1(S_A - S_B) \simeq (A_B/L_B)\tfrac{1}{2}a(T_1^2 - T_B^2)\tau_{AB} \qquad \ldots\ldots(2)$$

i.e.

$$T_1(S_A - S_B) \simeq \tfrac{1}{2}a(T_1^2 - T_B^2)\eta.$$

To obtain T_2, the temperature at which the working substance remains during the isothermal path C → D of the demagnetization, one notes that $S_B = S_C$, whence

$$H_B/T_1 = H_C/T_2 \qquad \ldots\ldots(3)$$

if the Brillouin function may be used to describe the behaviour of the paramagnetic working substance in the fields and at the temperatures of operation.

To calculate T_R the reservoir temperature one uses the fact that the heat flowing into the working substance during the isothermal path C → D of the demagnetization is equal to the sum of the heats flowing through the 'open' valve V_R and the 'closed' valve V_B, or symbolically

$$(A_R/L_R)\tfrac{1}{2}a(T_R^2 - T_2^2)\tau_{CD} + (A_B/L_B)\tfrac{1}{4}b(T_B^4 - T_2^4)\tau_{CD} = nT_2(S_D - S_C). \qquad \ldots\ldots(4)$$

In an ideal Carnot cycle $S_D - S_C$ would be equal to $S_A - S_B$. In practice the cycle is not ideal and, as in our previous work, one accounts for the lack of ideality by putting

$$S_D - S_C = \beta(S_A - S_B) \qquad \ldots\ldots(5)$$

where from practical experience the function β has a value of about 0·5.

Using equations (2), (3) and (5) and putting the ratio of the thermal valve sizes $r \equiv (A_R/L_R)/(A_B/L_B)$, one gets from equation (4) the following evaluation of T_R

$$T_R^2 = T_2^2 + \frac{1}{r}\left\{\frac{\beta H_C}{H_B}(T_1^2 - T_B^2)\left(\frac{\tau_{AB}}{\tau_{CD}}\right) - \frac{\tfrac{1}{4}b}{\tfrac{1}{2}a}(T_B^4 - T_2^4)\right\} \qquad \ldots\ldots(6)$$

which gives T_R as an implicit function of η, since by equations (2) and (3) T_1 and T_2 are functions of η. Figure 4 gives the results of this computation of T_R as a function of η and r for an engine using iron ammonium alum as the working substance and having as values of the fixed parameters $T_B = 1$°K, $H_A = 2$ kilooersteds, $H_B = 9$ kilooersteds, $H_C = 1{\cdot}5$ kilooersteds, $\beta = 0{\cdot}5$ and $\tau_{AB} = \tau_{CD}$.

To calculate Q_R^{ABCD} the net heat extracted from the reservoir during the complete cycle one assumes it to be equal to the difference between the heat flowing from the reservoir to the working substance during the isothermal demagnetization from C → D and the unwanted heat flowing back into the reservoir through the 'closed' valve V_R during the isothermal magnetization A → B. This assumption basically means that heat flows during the short period changes D → A and B → C are neglected. We therefore have

$$\begin{aligned} Q_R^{ABCD} &= (A_R/L_R)\tfrac{1}{2}a(T_R^2 - T_2^2)\tau_{CD} - (A_R/L_R)\tfrac{1}{4}b(T_1^4 - T_R^4)\tau_{AB} \\ &= r\eta n\{\tfrac{1}{2}a(T_R^2 - T_2^2)(\tau_{CD}/\tau_{AB}) - \tfrac{1}{4}b(T_1^4 - T_R^4)\}. \end{aligned} \qquad \ldots\ldots(7)$$

---

† This relation is approximate, since the heat flowing through the 'closed' thermal valve V_R is neglected. In fact with the values of the parameters used in practice, this approximation is accurate to a few per cent.

Figure 5 gives the results of numerical computation of Q_R^{ABCD}/n from equation (7) assuming, as before, iron ammonium alum for the working substance, $T_B = 1°K$, $H_A = 2$ kilooersteds, $H_B = 9$ kilooersteds, $H_C = 1{\cdot}5$ kilooersteds, $\beta = 0{\cdot}5$ and $\tau_{AB} = \tau_{CD}$. In the figure Q_R^{ABCD}/n is plotted as a function of η for various values of r. It is to be noted that as η is increased Q_R^{ABCD}/n rapidly falls. This can be understood from equation (7) in which the second term on the right-hand side represents the unwanted leakage of heat into the reservoir through the so-called 'closed' valve V_R during the period τ_{AB}. An increase of η increases the value of this term.

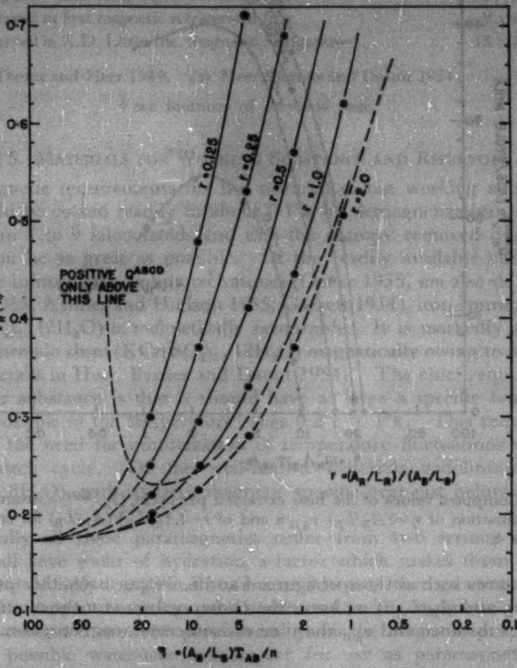


Figure 4. Computed values of the reservoir temperature T_R as a function of $\eta = (A_B/L_B)\,\tau_{AB}/n$ and of $r = (A_R/L_R)/(A_B/L_B)$, for conditions: $T_B = 1°K$, $H_A = 2$ kOe, $H_B = 9$ kOe, $H_C = 1{\cdot}5$ kOe, $a = 4$ w units, $b = 8 \times 10^{-2}$ w units, $\beta = 0{\cdot}5, \tau_{AB} = \tau_{CD}$, and iron ammonium alum as working substance.

The fact that, above certain values of η, Q_R^{ABCD} becomes zero also serves to limit the range of applicability of the family of curves of figure 4. In figure 4 the broken curve marks the boundary where Q_R^{ABCD} becomes zero. Only above this broken curve are the T_R values meaningful. It will be seen that this boundary curve is relatively narrow in range of η, so restricting severely the choice of valve sizes. It is to be noted that the sharp rise in the boundary curve for the low η

values is due to the fact that during the magnetization A → B the working substance would rise to relatively high temperatures, so raising also the value of T_2, whereas the rapid rise of the boundary curve for the high η values is due to the fact that with large η the working substance immediately after adiabatic demagnetization would warm up unduly rapidly.

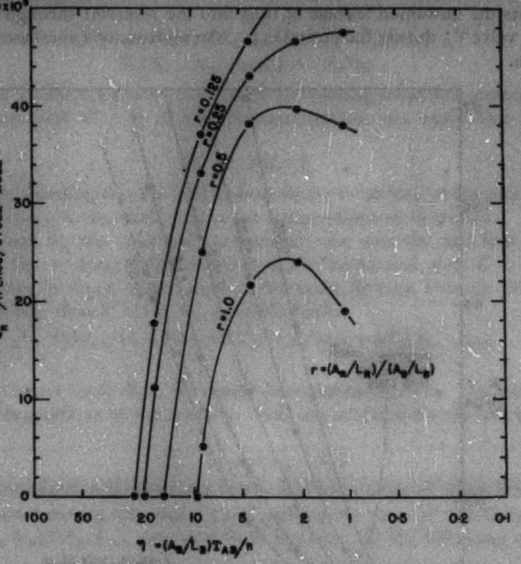


Figure 5. Computed values of the heat extracted per cycle, Q_R^{ABCD}/n, from the reservoir as a function of $\eta=(A_B/L_B)\ \tau_{AB}/n$ and of $r=(A_R/L_R)/(A_B/L_B)$ for conditions as in figure 4.

From curves such as those in figures 4 and 5, we can assess the optimum valve sizes A_B/L_B and A_R/L_R if we know the chosen values of n the number of moles of working substance and τ_{AB} the time of isothermal magnetization.

§ 4. SOME PERFORMANCE FIGURES

The first magnetic refrigerator made, which has already been described in detail (Heer, Barnes and Daunt 1954), used 15 g of iron ammonium alum as working substance, had $A_B/L_B = 10^{-2}$ cm and $\tau_{AB} = 43$ seconds. The value of η therefore was 13·8 cm sec mole^{-1}. A more recent model (Daunt *et al.* 1955) made by A. D. Little† used the following values: mass of iron ammonium alum 29 g, $A_B/L_B = 1{\cdot}8 \times 10^{-2}$ cm and $\tau_{AB} = 45$ seconds. η therefore was 13·6 cm sec mole^{-1}. The theoretical values of the net heat extraction per cycle Q_R^{ABCD}/n can be obtained from results as shown in figure 5. For systems, as used in practice, where $T_B = 1$°K, $H_A = 2$ kilooersteds, $H_B = 9$ kilooersteds, $H_C = 1\cdot5$ kilooersteds, $\tau_{AB} \simeq 45$ seconds and $r \simeq 0\cdot125$, the theoretical and observed values of Q_R^{ABCD}/n are given in table 1. It will be noted that the experimentally observed values are reasonably near the computed values.

---

† I am indebted to Dr. H. O. McMahon and Dr. I. Simon for these details and for permission to quote their results before publication.

Table 1. Net Heat Extraction Rate Q_R^{ABCD}/n in ergs/cycle per mole of Fe alum working substance at 0·40°K

Original estimate given in 1949 (a).	28×10^5
Computed value using eqn (7) of this paper.	27×10^5
Observed in first magnetic refrigerator (b).	9×10^5
Observed in A.D. Little Inc. magnetic refrigerator†.	18×10^5

(a) Daunt and Heer 1949. (b) Heer, Barnes and Daunt 1954.

† see footnote of previous page.

§ 5. MATERIALS FOR WORKING SUBSTANCE AND RESERVOIR

The magnetic requirements for the material of the working substance are that it should be cooled readily to about 0·1°K by demagnetization from initial fields of from 7 to 9 kilooersteds and that the entropy removed during initial magnetization be as great as possible. Of the readily available substances in common use in magnetic cooling techniques (Cooke 1955, see also de Klerk and Steenland 1955, Ambler and Hudson 1955, Garrett 1954), iron ammonium alum ($FeNH_4(SO_4)_2 . 12H_2O$) is magnetically favourable. It is markedly superior to potassium chromic alum ($KCr(SO_4)_2 . 12H_2O$) magnetically owing to its higher S value (see details in Heer, Barnes and Daunt 1954). The chief requirement for the reservoir substance is that it should have as large a specific heat per unit volume as possible in the temperature range 0·2°K to 1°K. This requirement is dictated by the need for minimization of temperature fluctuations during the course of each cycle. Of the well-known materials gadolinium sulphate ($Gd_2(SO_4)_3 . 8H_2O$), with its large magnetic specific heat just below 1°K, seems suitable.

Technically all these paramagnetics suffer from two serious drawbacks. First, they all have water of hydration, a factor which makes them chemically unstable, and secondly it is difficult to ensure adequate thermal contact between them and the copper mountings which are used in the magnetic refrigerator. We have investigated over the past several years the magnetic properties of a number of possible water-free substances† for use as paramagnetics at low temperatures and some preliminary findings on ferric acetylacetonate $(C_5H_7O_2)_3Fe$ (Daunt and Pillinger 1955) and on synthetic ruby (Corundum (Al_2O_3) with

---

† This work has been done in co-operation with Dr. W. L. Pillinger, Mr. J. W. Snider and Mr. K. Brugger and among the many substances investigated at very low temperature besides the two reported in this paper were chromic acetylacetonate, 'doped' corundum in the forms of 'Burma Sapphire' (Fe–Ti added) and 'green sapphire' (Sn–Va–Co added), spinel [$Mg(AlO_2)_2$] doped with Co–Ni–Fe (the so-called 'blue spinel' and aquamarine), various sintered aluminum oxide–chromium oxide systems as used for catalysts, and ordinary builders' brick. The latter, the use of which was suggested by Dr. I. Simon, contained 5% Fe_2O_3 by weight and showed a Curie law paramagnetic behaviour in the liquid helium temperature range. Indeed, we have already used brick for adiabatic demagnetization work with considerable success.

Cr_2O_3 impurity added) (Daunt, Snider and Brugger 1956, Heer and Rauch 1955) may be of interest.

Ferric acetylacetonate in compressed powder form follows a Curie–Weiss law in the liquid helium temperature range with a Weiss constant of 0·24°K. The ruby sample with small impurity content (sample No. I of tables 2 and 3) followed a Curie law. The Curie constants for these are given in table 2 together with those for some commonly used paramagnetic salts for comparison. Also given in table 2 are results obtained by Ambler and Hudson (1956) for ammonium hexafluorochromite, which also is water-free. From demagnetization experiments data on the entropy as a function of the magnetic temperature T^* were obtained also for $(C_5H_7O_2)_3Fe$ and ruby and from these findings the magnetic specific heat C^*, in the temperature range approximately 0·3°K to 1°K, where it is given by $C^* = a/T^{*2}$, could be computed. The observed values of CT^{*2} are given in table 2†. The values are given per cubic centimetre, since it is the volume which is the technically interesting parameter. Detailed presentations of the entropy diagrams are left to future publications; but in order to assess the final magnetic temperatures attainable on adiabatic demagnetization from initial magnetic fields of about 10 and 20 kilooersteds at 1°K, table 3 presents the values T_S^* corresponding to entropy values given by $S/R = \frac{1}{2} \ln(2S+1)$ and $S/R = \frac{1}{4} \ln(2S+1)$ respectively.

Table 2. Curie Constants and Magnetic Specific Heats per unit volume of some Paramagnetic Materials

Substance	Curie Constant × 10^3 (e.m.u. deg cm^{-3})	$C^*T^{*2} \times 10^{-3}$ (ergs deg^2 cm^{-3})
$KCr(SO_4)_2.12H_2O$	6·70	4·30
$(NH_4)_3CrF_6$	16·7	21·5
Ruby I	0·70	1·54
Ruby II	2·77	6·0 (calc)
$FeNH_4(SO_4)_2.12H_2O$	15·5	4·26
$Gd_2(SO_4)_3.8H_2O$	31·4	124
$(C_5H_7O_2)_3Fe$	17·8	163

As table 2 shows, $(C_5H_7O_2)_3Fe$ has a remarkably large magnetic specific heat below 1°K, even larger than gadolinium sulphate. It is therefore an ideal substance for use as reservoir material in the magnetic refrigerator. It is very stable chemically and can be readily cast in cold setting plastics without chemical reaction taking place. We have found that by casting the ferric acetylacetonate and a plastic mixture in a mould having copper strips running through it, a final assembly is obtained which is hard and durable and which maintains excellent thermal contact between the copper and the paramagnetic. The relatively high temperatures obtained on adiabatic demagnetization of $(C_5H_7O_2)_3Fe$, however (see table 3), preclude its use for a working substance in the magnetic refrigerator.

The two samples of synthetic ruby listed in tables 2 and 3 differed in Cr_2O_3 concentration. For both, however, the Curie constant per cubic centimetre was much smaller than that for either chromic or ferric alum. Moreover, it is considered that 'Ruby II' has the highest Cr_2O_3 addition ($\simeq 2\%$) possible in the production of such gems. The ruby therefore does not fulfil the requirement of large entropy changes *per unit volume* in the magnetic fields employed for the working substance of the magnetic refrigerator. It may, on the other hand, be useful as material for a sub-standard thermometer in the temperature range 0·1°K to 20°K, since it has two desirable qualities for that purpose, namely: (i) extreme chemical stability and (ii) practicability of gem-to-metal bonding (after firing on a metal coating on the ruby) to ensure good thermal contact. Investigations on the ruby for thermometric applications are now being carried out by us.

---

† Unfortunately equation (2) giving the specific heat of $(C_5H_7O_2)Fe$ in the paper by Daunt and Pillinger (1955) is in error. It should read $C^*/R = 0{\cdot}52/T^{*2}$ per mole.

Table 3. Magnetic Temperatures T_S^* of some new Paramagnetics as a Function of Entropy

Substance	$S/R=\frac{1}{2}\ln(2S+1)$	$S/R=\frac{1}{4}\ln(2S+1)$
Ruby I	0·162	0·095
$(NH_4)_3CrF_6$	0·18	0·11
$(C_5H_7O_2)Fe$	0·49 (0·45°K)	0·41 (0·33°K)

The ammonium hexafluorochromite, while possessing the quality desirable for a working substance of high Curie constant per cubic centimetre fails, as is evident from table 3, in not cooling to sufficiently low temperatures on demagnetization from fields of less than 10 kilooersteds. At present therefore we continue to use ferric alum for working substance, in spite of its technical inferiorities of having water of hydration, marked chemical reactivity, etc.

§ 6. USES OF THE MAGNETIC REFRIGERATOR

Some of the possible uses for the magnetic refrigerator are as follows:

(1) The production of isothermal enclosures at temperatures down to 0·25°K. This would allow accurate calorimetry to be performed in the range 0·25°K to 1°K, a range of temperature in which for example the study of electronic specific heats of metals would be of significant importance. In addition this would allow isothermal magnetic measurements to be made of many paramagnetic substances at or near their ferromagnetic or antiferromagnetic transition points with the consequent establishment of complete phase diagrams for the transformations.

(2) The establishment of a new low in isothermal temperature reservoirs to act as an 'initial' temperature for magnetic cooling. This would be of value also in the field of nuclear polarization.

(3) Provision of isothermal enclosures to 0·25°K to allow a reduction in the unwanted heat leaks to experimental arrangements cooled to the millidegree temperature region.

(4) The establishment of a constant temperature helium bath down to 0·25°K for the study, for example, of the properties of both liquid ^4He and ^3He.

The uses enumerated above for the magnetic refrigerator only represent a few selected examples of the many uses to which it may be put. In all, however, it must be emphasized that its primary feature is that it provides an *isothermal reservoir*, allowing exothermic experiments to be performed in the temperature range down to 0·25°K.

It should be noted also that a two-stage magnetic refrigerator is in principle possible and it is hoped that before long such a two-stage machine will be in operation to provide constant temperature reservoirs down to temperatures well below 0·25°K.

ACKNOWLEDGMENTS

I wish to acknowledge the co-operation of Drs. C. V. Heer and W. L. Pillinger and of Messrs. C. B. Barnes, J. W. Snider and K. Brugger, of the Low Temperature Physics Laboratory, Ohio State University, for the many contributions they have made to various phases of the work. I also wish to thank Dr. H. O. McMahon, Dr. I. Simon and Mr. J. Reitzel of A. D. Little Inc. for their valuable discussions and help and for their permission to quote their unpublished results. My thanks also go to the U.S. National Science Foundation for their continuing assistance in the later phases of the work.

REFERENCES

Ambler, E., and Hudson, R. P., 1955, *Rep. Progr. Phys.*, **18**, 251 (London: Physical Society); 1956, *Phys. Rev.*, **102**, 916.

Chase, C. E., and Herlin, M. A., 1955, *Phys. Rev.*, **97**, 1447.

Cooke, A. H., 1955, *Progress in Low Temperature Physics* (Amsterdam: North-Holland).

Collins, S. C., and Zimmerman, F. J., 1953, *Phys. Rev.*, **90**, 991.

Daunt, J. G., 1953, *Physica*, **19**, 787.

Daunt, J. G., and Heer, C. V., 1949 a, *Proceedings of the International Conference on Low Temperature Physics*, M.I.T., p. 65; 1949 b, *Phys. Rev.*, **76**, 985.

Daunt, J. G., Heer, C. V., McMahon, H. O., Reitzel, J., and Simon, I., 1955, *Bull. Inst. Internat. du Froid*, Annexe 1955–3, Supplement, p. 362.

Daunt, J. G., and Pillinger, W. L., 1955, *Bull. Inst. Internat. du Froid*, Annexe 1955–3, Supplement, p. 158.

Daunt, J. G., Snider, J. W., and Brugger, K., 1956, *Bull. Amer. Phys. Soc.*, **1**, 116.

Debye, P., 1926, *Ann. Phys., Lpz.*, **81**, 1154.

Garrett, C. G. B., 1954, *Magnetic Cooling* (New York: John Wiley).

Giauque, W. F., 1927, *J. Amer. Chem. Soc.*, **49**, 1870.

Giauque, W. F., and Macdougall, D. P., 1933, *Phys. Rev.*, **43**, 768.

Gorter, C. J., 1948, *Les phenomènes cryomagnétiques* (Paris: Collège de France), p. 76.

de Haas, W. J., and Rademakers, A., 1940, *Physica*, **7**, 992.

de Haas, W. J., Wiersma, E. C., and Kramers, H. A., 1933, *Physica*, **13**, 175; 1934, *Ibid.*, **1**, 1.

Heer, C. V., 1949, *Dissertation*, The Ohio State University.

Heer, C. V., Barnes, C. B., and Daunt, J. G., 1953, *Phys. Rev.*, **91**, 412; 1954, *Rev. Sci. Instrum.*, **25**, 1088.

Heer, C. V., and Daunt, J. G., 1949, *Phys. Rev.*, **76**, 854.

Heer, C. V., and Rauch, R. J., 1955, *Bull. Inst. Internat. du Froid*, Annexe 1955–3, Supplement, p. 218.

Hwang, C. P., Fulton, C. D., and Fairbank, W. M., 1956, *Bull. Amer. Phys. Soc.*, **1**, 217.

de Klerk, D., and Steenland, M. J., 1955, *Progress in Low Temperature Physics* (Amsterdam: North-Holland).

Mendelssohn, K., and Olsen, J. L., 1950, *Proc. Phys. Soc.* A, **63**, 1182.

Mendelssohn, K., and Rosenberg, H. M., 1952, *Proc. Phys. Soc.* A, **65**, 388; 1953, *Proc. Roy. Soc.* A, **218**, 190.

Nicol, J., and Daunt, J. G., 1952, *Proceedings of the Cryogenics Conference*, Schenectady. p. 68.

Olsen, J. L., and Renton, C. A., 1952, *Phil. Mag.*, **43**, 946.

Osborne, D. V., 1956, *Phil. Mag.*, **1**, 301.

Rademakers, A., 1949, *Physica*, **15**, 849.

Oblique Ray Paths in the Ionosphere

BY J. HASELGROVE[†]
The University Mathematical Laboratory, Cambridge

Communicated by J. A. Ratcliffe ; MS. received 25th October 1956,
and in final form 22nd March 1957

Abstract. An electronic digital computer has been used to calculate ray paths in a parabolic ionospheric layer when the waves are incident in the plane of the magnetic meridian. This is a special application of a very general method. The results of the calculations are compared with what would be deduced from vertical soundings by the standard approximations of using Martyn's and Breit and Tuve's theorems as if there were no field.

§ 1. INTRODUCTION

THE path taken in the ionosphere by radio energy incident obliquely upon it is important in connection with short-wave radio communication. The calculation of such a path, even in a horizontally stratified layer, is made extremely difficult by the effect of the earth's magnetic field. Various graphical and numerical methods have been given (for example, by Poeverlein 1949, 1950, and Millington 1954) for determining ray paths; in this paper another method, using an electronic digital computer, is described. Some results are given of calculations performed by the computer EDSAC at the Cambridge University Mathematical Laboratory.

The two quantities which are of interest in a ray path are the horizontal distance d between the sender and the point at which the reflected energy returns to the earth, and the equivalent path P' (which is defined as the 'time of flight' of a pulse from its transmission to its reception, multiplied by the velocity of waves in free space). For a given ionosphere d and P' can be regarded as functions of the angle of incidence i and the frequency f. If there were no magnetic field, P' and d for any i and f could be determined, from measurements with vertical transmission ($i=0$) and appropriate frequency, by the following well-known relations:

(Martyn's theorem) $$P'(i,f)=\sec i\ P'(0,f\cos i), \qquad \ldots\ldots(1)$$

(Breit and Tuve's theorem) $$d=P'\sin i. \qquad \ldots\ldots(2)$$

A combination of these two equations gives

$$d(i,f)=\tan i\ P'(0,f\cos i). \qquad \ldots\ldots(3)$$

Equations (1)–(3) are independent of the shape of the ionospheric layer, provided that it is horizontally stratified and that the curvature of the earth is neglected. In the presence of a magnetic field, however, Martyn's and Breit and Tuve's theorems are not valid and there is no simple relation, holding for all layers, between the quantities for vertical and oblique rays. But since calculations concerning propagation in the presence of a magnetic field have so far been very long and laborious, the behaviour of oblique rays in the ionosphere has been predicted approximately from vertical soundings of the ionosphere by using equations (1)–(3) as if there were no field. The purpose of this paper is to show the errors introduced when this approximation is made, for the ordinary ray in a parabolic layer over a flat earth.

[†] Present address: 35 Belfield Road, Didsbury, Manchester 20.

The calculations are used in the following way. For each oblique ray path with frequency f and angle of incidence i the equivalent path P' and the horizontal range d are found. The equivalent height $\frac{1}{2}P'(0,f\cos i)$ of a vertical ray of frequency $f\cos i$ in the same ionospheric layer is then calculated, and the quantities

$$P_{\mathrm{v}}'(i,f)=\sec i\, P'(0,f\cos i)$$

$$d_{\mathrm{v}}(i,f)=\tan i\, P'(0,f\cos i)$$

are formed. The subscript 'v' is used to show the derivation from a vertical ray calculation. The quantities P_{v}' and d_{v} are those which would be given by an observer who found a value of $P'(0,f\cos i)$ experimentally and then used Martyn's and Breit and Tuve's theorems to find approximations for the range and equivalent path of an oblique ray. If there were no field the pairs of quantities P', P_{v}' and d, d_{v} would be equal. If there is a field these pairs of quantities are not equal and a comparison of them gives a measure of the error introduced by using Martyn's and Breit and Tuve's theorems. This comparison is made in § 4 for the calculations which have been performed.

Calculated values of horizontal range $d(i,f)$ can be used in two ways. For a fixed frequency f and a series of angles i the smallest distance S or 'skip distance' can be found, or for a fixed distance d_0 a series of pairs of values f and i which make $d=d_0$ can be found, so that of these the greatest value of f, the maximum usable frequency, is determined. The maximum usable frequency can also be found more simply by plotting a series of skip distances for different frequencies in the same ionosphere, as in figure 5. The maximum usable frequency is then merely read off against the required distance. Calculations of this type are much used in radio communication engineering.

The method of calculating ray paths described in this paper can in principle take into account all the variable parameters which occur. For a complete study of the problem these would be the shape, thickness and height of the ionized layer, strength and direction of field, direction of incident ray, collision frequency and, in addition, possible gradients of these quantities in a horizontal plane. In practice, however, limits are imposed by the capacity of the computer and by the time available. Thus the results presented here are considerably restricted, though it is hoped to widen the scope when a larger and faster machine is available.

In this paper only ordinary rays are considered. Propagation is in the North–South direction so that the ray paths lie in a plane which also contains the magnetic field vector. The ionospheric layer is taken to be horizontally stratified and parabolic and the magnetic field to be constant in strength and direction throughout the layer. Although a parabolic layer is taken for convenience it is not necessary to use an analytic formula and it would be possible, for instance, to use an experimentally determined profile in tabular form and so compare the results directly with experience. The curvature of the earth is neglected; this makes practically no difference to the time of computation but reduces the number of variable parameters to be considered, by removing the effect of the height and thickness of the layer (except that the ratio of these two quantities enters into skip distance calculations even with a flat earth). The effect of collisions is neglected. Thus there remain as variable parameters the strength of the magnetic field, the dip angle, the critical frequency of the layer and the frequency and angle of incidence of the ray.

The method of calculating ray paths is described in § 2, and in § 3 the effect of a magnetic field on some ray paths is shown. § 4 is concerned with the inaccuracy of Martyn's theorem, and § 5 with the effect of this inaccuracy on the calculation of horizontal ranges and skip distances.

§ 2. THE METHOD OF COMPUTATION

2.1. *The Differential Equations*

In addition to the symbols defined already, the following notation will be used: f_0 plasma frequency at a point in the ionosphere, $X=f_0^2/f^2=Ne^2/\epsilon_0\pi mf^2$ (N, e, ϵ_0 and m in the usual notation), X_m maximum value of X for the layer, so that $X_m=f_c^2/f^2$ where f_c is the critical frequency of the layer, f_H gyro-frequency, $Y=f_H/f$, μ refractive index, ϕ dip angle, ψ angle between the wave normal and the magnetic field vector.

The Appleton–Hartree formula gives μ in terms of X, Y, and ψ. The frequency does not enter explicitly into any of the formulae, so that if the ionized layer is taken to be parabolic the layer, magnetic field and frequency are sufficiently defined by the three quantities X_m, Y and ϕ. The electron density at a height h measured from the bottom of a parabolic layer is given by the formula

$$X=X_m\{1-(1-h/h_t)^2\},$$

where h_t is the half-width. The value of the half-width is immaterial since, except for skip distances (§ 5), the results depend linearly upon it; throughout this paper the half-width is taken to be 100 km.

At any point on the ray the following quantities must be known: x, y, the horizontal and vertical coordinates, and χ, the angle between the horizontal and the wave normal (thus $\chi=\phi+\psi$).

For brevity a subscript is used to indicate partial differentation, thus:

$$\mu_x\equiv\left(\frac{\partial\mu}{\partial x}\right)_{y,\chi}, \qquad \mu_y\equiv\left(\frac{\partial\mu}{\partial y}\right)_{x,\chi}$$
$$\mu_\psi\equiv\left(\frac{\partial\mu}{\partial\psi}\right)_{X,Y}=\left(\frac{\partial\mu}{\partial\chi}\right)_{x,y}, \qquad \mu_f\equiv\left(\frac{\partial\mu}{\partial f}\right)_{x,y,\chi},$$
$$\mu_X\equiv\left(\frac{\partial\mu}{\partial X}\right)_{Y,\psi}, \qquad \mu_Y\equiv\left(\frac{\partial\mu}{\partial Y}\right)_{X,\psi}.$$

For the equivalent path the group refractive index μ_G is required; it is given by

$$\mu_G=\mu+f\mu_f=\mu-2X\mu_X-Y\mu_Y.$$

The ray path and equivalent path are calculated together by a modified form of the Runge–Kutta method of solving differential equations. A programme for this type of calculation is available in the EDSAC library of sub-routines (Gill 1951). A set of simultaneous first-order differential equations is used, the dependent variables being x, y, χ and P'. The equations are given here without proof; the derivations may be found in a previous paper (Haselgrove 1954, § 3.1). The value of the independent variable t nowhere enters into the calculation; t is in fact defined by the relation

$$\frac{dt}{ds}=\mu^2(\mu^2+\mu_\psi^2)^{-1/2},$$

s being the path length. The equations are

$$\frac{dx}{dt} = \frac{1}{\mu^2}(\mu \cos \chi - \mu_\chi \sin \chi),$$

$$\frac{dy}{dt} = \frac{1}{\mu^2}(\mu \sin \chi - \mu_\chi \cos \chi),$$

$$\frac{d\chi}{dt} = \frac{1}{\mu^2}(\mu_y \cos \chi - \mu_x \sin \chi),$$

$$\frac{dP'}{dt} = \frac{1}{\mu}\mu_G.$$

2.2. *General Description of the Programme*

The programme written for the EDSAC deals with the ray in the ionosphere only, the path in free space below the layer being added later if required. In using the EDSAC, the programme for either oblique or vertical rays is read into the machine, which then proceeds to read the initial conditions for a ray and perform the calculation. When one ray is finished the initial conditions for the next are read, and so on. These initial conditions consist of values for X_m, Y, ϕ and i. At the start of an oblique ray x, y and P' are set equal to zero and χ is set to $(\frac{1}{2}\pi - i)$. The same sets of numbers are used for both oblique and vertical rays; in the latter case the frequency must be multiplied by a factor $\cos i$ (see § 1), and this is effectively done by the programme, which replaces X_m by $X_m/\cos^2 i$, and Y by $Y/\cos i$.

It should be remarked that the ray paths are reciprocal, so that it is only necessary to consider propagation in one sense, say from South to North in the northern hemisphere. This is because the ray direction at any height has only two possible values (these can be calculated from the wave normal direction, itself given by Snell's law). The system in fact satisfies the conditions of Budden's reciprocity theorem (Budden 1954), though the ray path problem is only a special case of that theorem.

An oblique ray takes from 2 to 20 minutes to compute, the majority taking about 7 minutes. Values of the variables are printed at intervals along the ray so that the complete path is available. A further quantity $\mu \cos \chi$, which by Snell's law should remain constant, is also printed at intervals and provides a good check on the accuracy of the process and of the arithmetic. A vertical ray takes about $2\frac{1}{2}$ minutes to compute, the method being similar to that used by Shinn (1953).

The accuracy obtained can be greater than is needed for practical purposes; all the results used in this paper were found to within 0·1 km and most were much more accurate. The most difficult rays to compute are those with a small angle of incidence, for near the top of the ray where the curvature is large the numbers change rapidly and if they are not to exceed the capacity of the machine at one point they are likely to become insignificantly small at another. The limiting case of this is when the ray has a cusp (Poeverlein 1950), in which case the direct method cannot be used. The difficulty of keeping the numbers within the capacity of the machine would be much lessened if floating-point arithmetic were used, but with the EDSAC this would increase the time of calculation by a factor of about 30 and is therefore impracticable. The lower limit for i in the present calculations has been in the region of 30° or 40°, depending on the strength of the field and the dip angle.

§ 3. SOME RAY TRAJECTORIES

A number of ray paths were computed for parabolic layers with values of X_m equal to 2 and $\frac{1}{2}$. For each layer various magnetic field strengths were taken, given by the values $Y=\frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 4$ and infinity. The infinite case was added to show the limit for increasing Y, and was found by taking the limiting form of the Appleton–Hartree formula as Y tends to infinity. For each layer and field strength, dip angles of 0°, $22\frac{1}{2}$°, 45°, $67\frac{1}{2}$° and 90° were taken. The equivalent paths and horizontal ranges for the case $Y=0$ were found analytically.

The results are shown for the ordinary ray in figures 1, 2 and 3. They all refer to the ray path in the ionosphere only, and to a layer with half-width 100 km. In interpreting the curves it should be remembered that $X_m = f_c^2/f^2$ and $Y=f_H/f$. In practical applications, the fixed quantity is f_H, the gyro-frequency, and the wave frequency f is given by $f=f_H/Y$. The critical frequency f_c of the layer is then given by

$$f_c = fX_m^{1/2} = f_H X_m^{1/2}/Y.$$

In figure 1 some ray paths are drawn, all for rays incident at an angle $i=60°$ on a layer with $X_m=2$. The imposed magnetic field is given three magnitudes corresponding to the values $Y=\frac{1}{2}$, 2 and ∞, and a series of dip angles ϕ. The vertical field ($\phi=90°$) has very little effect on the ray, but the horizontal field increases the range by a factor of nearly 3 when $Y=2$. The point R on the horizontal axis at a distance corresponding to 45·3 km gives the horizontal range

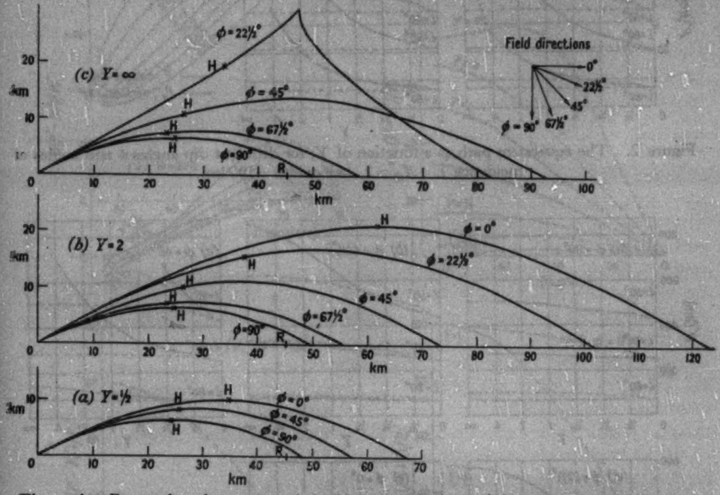


Figure 1. Ray paths when a wave is incident in the plane of the magnetic meridian at an angle $i=60°$ on a parabolic layer of half-width 100 km, with critical frequency $=\sqrt{2}\times$ wave frequency (i.e. $X_m=2$). The imposed magnetic field has magnitude given by $Y=f_H/f$ and dip angle ϕ. The wave normals are horizontal at the points marked H. The point R would be reached by the ray if there were no field.

of the ray when there is no field; the height of reflection when there is no field is the same as for $\phi=90°$, $Y\neq0$. The point H on each ray shows where the wave normal is horizontal; for $\phi=0°$ and $\phi=90°$ this is at the point where the ray is also horizontal. One ray with a cusp is shown; the two halves of this were calculated separately, starting from the bottom of the ionosphere.

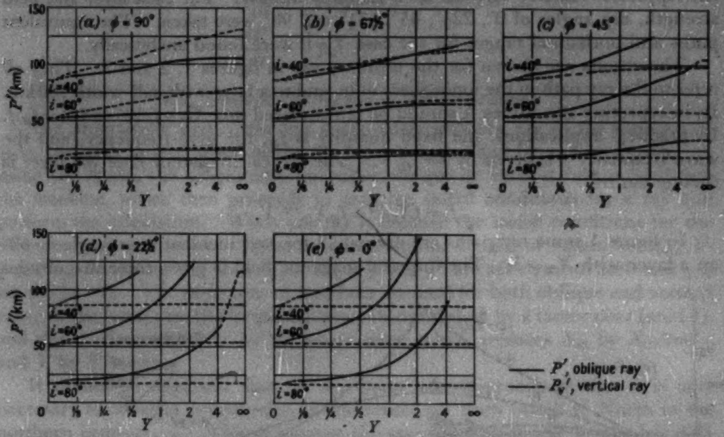


Figure 2. The equivalent path as a function of Y, for different dip angles ϕ and angles of incidence i. $X_m=2$, half-width=100 km.

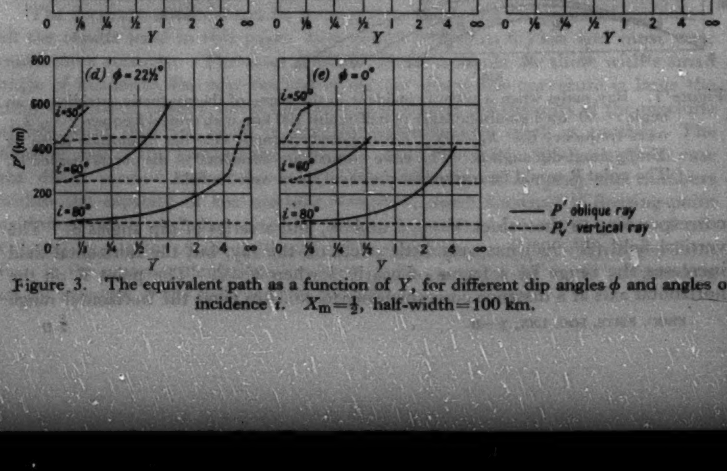


Figure 3. The equivalent path as a function of Y, for different dip angles ϕ and angles of incidence i. $X_m=\frac{1}{2}$, half-width=100 km.

In figure 2 the full curves show how the equivalent path varies with Y, for a layer with $X_{\rm m} = 2$ and for different dip angles. In each case angles of incidence $i = 80°$, $60°$ and $40°$ were taken. Again it is seen that the equivalent path is altered most when the magnetic field is horizontal.

In figure 3 the full curves show the same thing for a layer with $X_{\rm m} = \frac{1}{2}$, taking $i = 80°$, $60°$ and $50°$. In this case some of the rays penetrate the ionosphere and one effect of the field is to increase the angle of incidence at which reflection just ceases to occur. Since P' becomes infinite at this critical angle, the effect of the field on P' is very marked in this neighbourhood, as is shown by the curves for $i = 50°$.

§ 4. MARTYN'S THEOREM

In figures 2 and 3 the broken curves show how the quantity $P_{\rm V}'$, calculated as described in § 1, varies with Y. These broken curves meet the full curves on

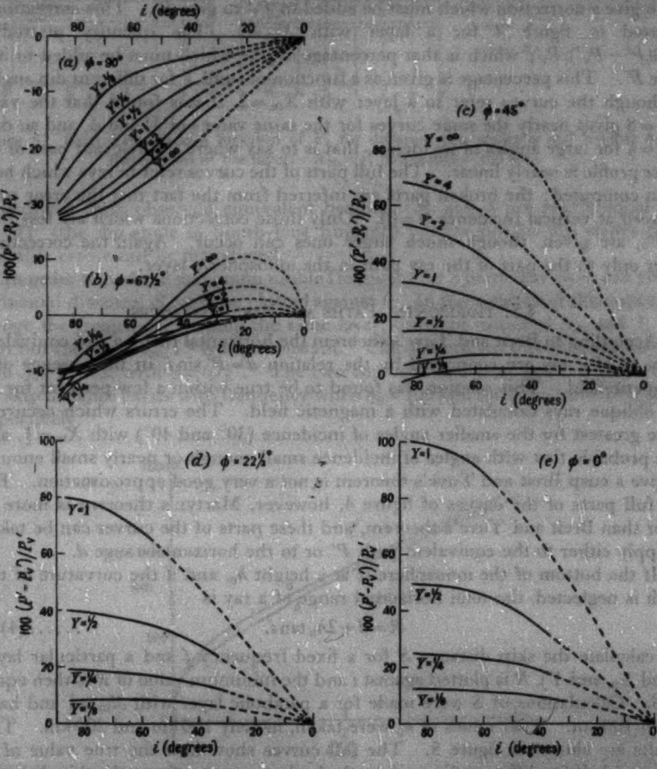


Figure 4. Percentage error introduced by using Martyn's theorem to derive equivalent paths. The quantity plotted is the percentage of $P_{\rm V}'$ which must be added to $P_{\rm V}'$ to give P'.

the left-hand axes ($Y=0$). The horizontal field has no effect on P_V', since P_V' is calculated from a vertical ray path for which propagation is then transverse, but the vertical field has more effect than on the oblique rays.

As described in § 1, it is customary in radio-communication engineering to assume that P_V', calculated from values of the equivalent height observed at vertical incidence, represents the true value of the equivalent path for oblique incidence. In fact this assumption is not justified, since the true value is P'. A measure of the error introduced when the assumption is made is given by the distance between the full and broken curves in figures 2 and 3, which represents the difference between the values of P_V' and P'. The error can become very large, as for instance in figure 2 for $\phi=0°$, but it should be remembered that these figures refer to the path in the layer only, and that the effect will be lessened by the addition of the path below the layer.

Another way of representing the error introduced by using P_V' instead of P' is to give a correction which must be added to P_V' to give P'. This correction is plotted in figure 4 for a layer with $X_m=2$. The quantity plotted is $100(P'-P_V')/P_V'$, which is that percentage of P_V' which must be added to it to give P'. This percentage is given as a function of i and Y for different dip angles. Although the curves refer to a layer with $X_m=2$, it was found that the value $X_m=8$ gives nearly the same curves for the same values of Y and ϕ, and so does $X_m=\frac{1}{2}$ for large angles of incidence, that is to say where the relevant part of the layer profile is nearly linear. The full parts of the curves refer to rays which have been computed; the broken parts are inferred from the fact that the error must be zero at vertical incidence ($i=0$). Only those corrections which are less than 100% are given, though much larger ones can occur. Again the corrections refer only to the part of the ray path in the ionospheric layer.

§ 5. HORIZONTAL PATHS AND SKIP DISTANCES

According to Breit and Tuve's theorem the horizontal range d and equivalent path P' of a ray are connected by the relation $d=P'\sin i$, in the absence of a magnetic field. This relation was found to be true within a few per cent for all the oblique rays calculated with a magnetic field. The errors which occurred were greatest for the smaller angles of incidence (30° and 40°) with $X_m=\frac{1}{2}$, and it is probable that with angles of incidence small enough or nearly small enough to give a cusp Breit and Tuve's theorem is not a very good approximation. For the full parts of the curves of figure 4, however, Martyn's theorem is more in error than Breit and Tuve's theorem, and these parts of the curves can be taken to apply either to the equivalent path P' or to the horizontal range d.

If the bottom of the ionosphere is at a height h_0, and if the curvature of the earth is neglected, the total horizontal range of a ray is

$$R=d+2h_0\tan i. \qquad \ldots\ldots(4)$$

To calculate the skip distance S for a fixed frequency f and a particular layer (fixed X_m and Y), R is plotted against i and the minimum value of R is then equal to S. Calculations of S were made for a parabolic layer with $X_m=\frac{1}{2}$ and half-width 100 km. Two values of h_0 were taken, namely 100 km and 200 km. The results are shown in figure 5. The full curves show how the true value of S varies with Y for different dip angles; the broken curves refer to the skip distance S_V which is found by using d_V instead of d in (4). S_V is the quantity which would

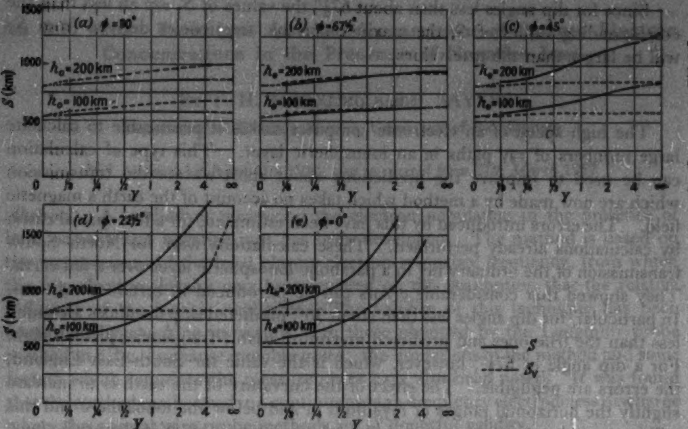


Figure 5. Skip distance as a function of Y, for different dip angles ϕ and heights h_0 of the bottom of the layer. $X_m=\frac{1}{2}$, half-width=100 km.

be deduced from observed equivalent heights. S_v and S are nearly equal for $\phi=67\frac{1}{2}°$ (the dip angle in South-East England) but for smaller dip angles considerable errors can be introduced.

In order to find the maximum usable frequency for a particular layer and given horizontal distance, S must be plotted against f. In the notation of this paper, to change the frequency, keeping the same layer, different values of X_m and Y are used but the ratio X_m/Y^2 is kept constant. Calculations have been made for the ratio $X_m/Y^2=2$, with $\phi=67\frac{1}{2}°$. The two curves A and B of figure 6 show how the skip distance varies with frequency when $h_0=100$ km, for $Y^2=0$ and $Y^2=\frac{1}{2}X_m$ respectively.

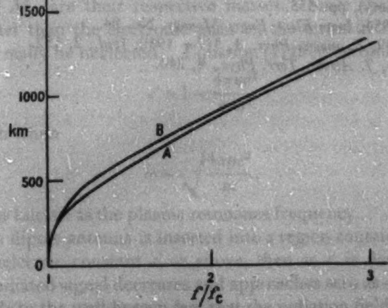


Figure 6. Skip distance as a function of frequency. Half-width=100 km, $h_0=100$ km. A, $Y^2=0$; B, $Y^2=\frac{1}{2}X_m$, $\phi=67\frac{1}{2}°$.

Since for dip angles less than about 67½° the values of S_v are all less than the corresponding values of S, the maximum usable frequencies derived from S_v will be larger than the true values.

§ 6. CONCLUSION

The high speed of an electronic computer makes it practicable to calculate large numbers of ray paths in an ionospheric layer. This type of calculation can be used to improve the predictions about short-wave radio transmission which are now made by a method which takes no account of the earth's magnetic field. The errors introduced by this have been estimated, for a few special cases, by calculations already performed. These calculations were for North–South transmission of the ordinary ray in a parabolic ionospheric layer over a flat earth. They showed that considerable errors can be introduced in certain conditions. In particular, for dip angles less than about 30°, predicted skip distances are much less than the true ones and predicted maximum usable frequencies are too high. For a dip angle of 67½°, however, which is the value for South-East England, the errors are negligible. The effect of the curvature of the earth is to increase slightly the horizontal ranges of rays both in and below the ionosphere and this will increase the error unless an appropriate correction is made in the predictions. It is not known why the error given by these calculations is of the opposite sign to that found in practice, when predicted maximum usable frequencies are always too low, but it is hoped that further calculations with other layer profiles and different directions of propagation may throw some light on the matter.

ACKNOWLEDGMENTS

I am indebted to the Director and staff of the University Mathematical Laboratory, Cambridge, for permission to use the EDSAC and to Dr P. C. Clemmow for his help in the course of supervising this work.

REFERENCES

BUDDEN, K. G., 1954, *Proc. Camb. Phil. Soc.*, **50**, 604.
GILL, S., 1951, *Proc. Camb. Phil. Soc.*, **47**, 96.
HASELGROVE, J., 1954, *Report of Conference on the Physics of the Ionosphere* (London : Physical Society), pp. 355.
MILLINGTON, G., 1954, *Instn Elect. Engrs Monogr.* No. 94.
POEVERLEIN, H., 1949, *Z. angew. Phys.*, **1**, 517 ; 1950, *Ibid.*, **2**, 152.
SHINN, D. H., 1953, *J. Atmos. Terr. Phys.*, **4**, 240.

An R.F. Probe Technique for the Measurement of Plasma Electron Concentrations in the Presence of Negative Ions

BY T. H. Y. YEUNG AND J. SAYERS
Department of Electron Physics, University of Birmingham

MS. received 1st November 1956, and in revised form 27th March 1957

Abstract. A method of measuring the electron population in the presence of both positive and negative ions has been developed. The method is based on the measurement of the local plasma electron resonance frequency, from which the free electron density may be calculated on the assumption that the relative influence of the ions is negligible. The validity of this technique has been confirmed by checking its results with those obtained by the Langmuir probe method in a d.c. discharge in noble gases; the limitations of the method and some factors which influence the measurements have also been noted. It was found that the method can be used in a continuous high-frequency electrodeless discharge where the normal wire probe methods are of doubtful validity.

The technique has been applied to a study of the decay of electron density in the afterglow of a pulsed discharge in iodine vapour; the cross section of electron attachment σ was found to be $7{\cdot}5 \pm 0{\cdot}5 \times 10^{-18}$ cm^2 for electrons at an estimated temperature of 8000°K.

§ 1. PRINCIPLES OF THE METHOD

In an ionized medium the real part of the dielectric constant ϵ' is depressed by the presence of the free charges in the medium. If we neglect the effect of collisions and there is no magnetic field, the equation for the dielectric constant is given by

$$\epsilon' = 1 - \frac{4\pi e^2}{\omega^2}\left(\frac{n}{m} + \frac{N_-}{M_-} + \frac{N_+}{M_+}\right) \qquad \ldots\ldots(1)$$

where n, N_- and N_+ are electron, negative ion and positive ion concentrations, and m, M_- and M_+ are their respective masses. Since the ionic mass M is very much greater than the electronic mass m, the terms in N_- and N_+ in the equation can normally be neglected. Equation (1) then becomes

$$\epsilon' = 1 - \frac{4\pi n e^2}{m\omega^2} \qquad \ldots\ldots(2)$$

and ϵ' will vanish when

$$\omega = \sqrt{\frac{4\pi n e^2}{m}}. \qquad \ldots\ldots(3)$$

This value of ω is known as the plasma resonance frequency.

If a Hertzian dipole antenna is inserted into a region containing free electrons and having a dielectric constant ϵ' as above, then as ϵ' is decreased the group velocity of the radiated signal decreases and approaches zero as ϵ' approaches zero. This corresponds to the well-known fact that the radiation field strength vanishes as ϵ' approaches zero.

There is, however, a second contribution to the field strength near the dipole which we may call the induction field and which varies as the inverse cube of the distance and which exceeds the radiation field at distances short compared with the wavelength. This field can be calculated using the quasi-stationary approximation† and in polar coordinates we have

$$E_r=\frac{2M\cos\theta}{4\pi\epsilon r^3}, \qquad E_\theta=\frac{M\sin\theta}{4\pi\epsilon r^3}, \qquad E_\phi=0$$

where M is the electric moment of the dipole. We therefore see that the induction electric field for a dipole of given electric moment varies inversely as the dielectric constant. In practical cases this does not lead to an infinite value for the field strength at plasma resonance because of the electron collision damping which has been neglected in the expression for ϵ' above. Radiation from the volume of resonant plasma is another factor limiting the field strength. (The corresponding circuit case is the series excited resonant circuit whose amplitude for a given impressed voltage is limited only by resistance and radiation losses.)

A maximum in the local induction field is therefore obtained at plasma resonance and in the experiments now described the location of this maximum is used as an indication of plasma resonance.

In the experimental apparatus two small probes in the form of coaxial h.f. transmission lines are introduced through opposite walls of the containing vessel into the region of the plasma to be explored. A few millimetres of the inner conductor of each is uncovered to act as an antenna and each probe is covered by a thin wall glass tube. One of the probes is connected to a source of microwave signal of small power output, and the other leads to a fairly sensitive receiver. Although the transmitting probe is not in the form of a symmetrical dipole it must still produce a dipole field and the analysis above must therefore apply. The receiving probe will detect a signal proportional to the field strength in its neighbourhood. Provided the distance between the probe ends is small compared with the wavelength of the probing signal it is found that when the plasma resonance frequency coincides with the probing signal frequency a maximum signal strength is transmitted to the receiving probe. When the plasma resonance frequency has been determined the electron density may be calculated from equation (3).

Since there is nothing in common between this method of measuring plasma electron densities and the Langmuir probe method, it was decided to carry out a cross check by using both methods in the same discharge in noble gases where, with the absence of negative ions, both methods are equally applicable.

§ 2. DISCHARGE TUBE AND ELECTRICAL CIRCUIT

All tubes were made of Pyrex about 6·5 cm in diameter and 20 to 25 cm in length. For the microwave probes two glass tubes of less than 1 cm in diameter were inserted into the centre of the discharge tube with 1 to 2 cm between the sealed ends, and lying diametrically opposite each other, as illustrated in figure 1 (*a*). For the experiments with d.c. discharges an oxide-coated cathode was used with a cylindrical anode and the positive column region of low voltage arcs was used for electron density measurements. For the work with h.f. excited discharges, tubes were constructed with external electrodes only. A block diagram of the experimental arrangement for the h.f. discharges is shown in figure 1 (*b*). The apparatus included three main sections: the transmitter, the receiver and the source of ionization. The transmitter was a microwave oscillator whose frequency was of the order of 300 Mc/s. For the receiver section, a simple crystal detector was used in d.c. discharges and c.w. high-frequency electrodeless discharges, but in the case of discharge afterglow studies a sensitive superheterodyne microwave receiver, with a triode converter and two-stage amplification was used. The source of ionization was a high tension d.c. regulated power supply for d.c. hot cathode discharges, a 700 kc/s power oscillator for h.f. electrodeless discharges in iodine vapour, and a two-microsecond pulsed power oscillator operating at 200 Mc/s and 50 c/s repetition frequency for the study of the electron attachment in iodine vapour.

---

† See, for example, *Antennae*, by J. Aharoni, Oxford University Press, 1946.

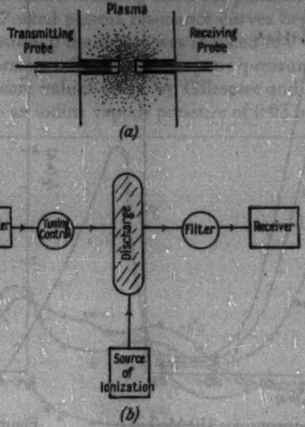


Figure 1.

§ 3. EXPERIMENTAL RESULTS

3.1. *D.C. Discharges and Comparison with Langmuir Probe Methods*

In d.c. discharges, the resonance curves were obtained by plotting the current of the crystal detector in the receiving section against the discharge current. Figure 2 shows a set of resonance curves for different powers from the main oscillator, where I_0 was the detector current obtained without the presence of a discharge. All resonance curves are of the same shape and have the following features: first there is a minimum galvanometer reading which is less than that obtained through vacuum, then the reading increases rapidly up to the resonant peak, and finally it decreases smoothly.

For an *LC* resonant circuit, the sharpness of resonance depends on the ohmic loss in the circuit. In the case of plasma resonance, collisions with neutral particles will destroy part of the directed momentum of the vibrating electrons. Since the collisional frequency of the electrons is proportional to the number of neutral particles present, the breadth of the resonance curve is pressure dependent. This pressure broadening of the plasma resonance curve is shown in figure 3, in which the relation between detector signal and d.c. discharge current is plotted for four different pressures.

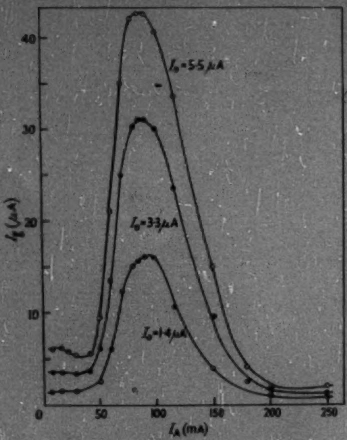


Figure 2. Tube II, neon; $f=230$ Mc/s, $p=0{\cdot}038$ mm, $n=6{\cdot}8\times10^{8}$ cm^{-3}.

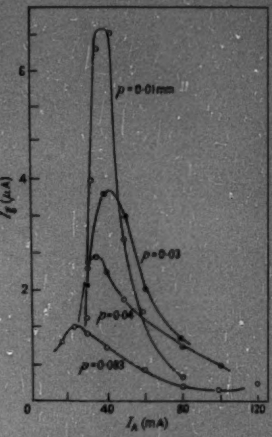


Figure 3. Tube III, neon; $f=215$ Mc/s, $I_e=0{\cdot}23$ μA.

For the purpose of the cross check between the microwave resonance and the Langmuir probe methods of measuring electron density the pressure was fixed at the region of 0·04 mm Hg to ensure a reasonably sharp plasma resonance. The d.c. discharge current was varied from zero to 300 milliamperes to provide a convenient range of plasma electron densities. The table shows the results of

	Tube II Neon			Tube III Neon						Tube III Argon	
N_{RF}	8·5	8·5	8·5	5·9	5·9	5·9	5·9	10	7·5	8·5	11
N_p	9·5	8·6	14	6·9	5·7	6·0	5·2	12	9	9	12
P	150	77	21	126	33	23	21	21	21	32	32
I_A	32	78	126	20	38	50	45	70	52	60	78
V_A	40·5	36·5	56	30	31	35	34	40	38	32	34

N_{RF}=electron density by r.f. probe technique (10^8 electrons cm^{-3}); N_p=electron density by Langmuir probe method (10^8 electrons cm^{-3}); P=pressure (microns); I_A=d.c. arc current (mA); V_A=d.c. arc voltage.

electron density measurements by the two methods for the positive column of low voltage arcs in neon and in argon in the 0·05 mm pressure range for various currents. The agreement evident here is within the expected range of experimental error.

3.2. *H.F. Electrodeless Discharge*

As a preliminary to electron attachment studies in iodine afterglows electron density measurements were made in steady h.f. discharges in iodine vapour. The iodine vapour was controlled by regulating the condensation temperature in the solid iodine reservoir which formed an appendage to the main tube. With the reservoir and tube at room temperature (about 22°c) the vapour pressure was so high that electron collision damping made the plasma resonance frequency impossible to identify. Peaked plasma resonance curves began to appear when the condensation temperature of the iodine was reduced to 0°c or less.

To determine the corresponding iodine vapour pressure the condensation temperature/vapour pressure values given by Gilespiet and Fraser (1936) were used; 0°c corresponds to an iodine vapour pressure of 0·03 mm Hg.

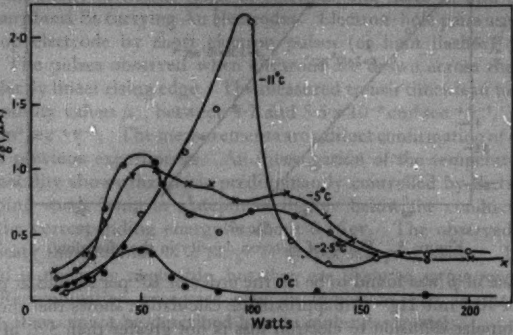


Figure 4. H.F. electrodeless discharge, iodine; $\lambda=140$ cm, $f=215$ Mc/s, $I_0=0{\cdot}06$ μA.

Figure 4 shows some of the curves, over a temperature range from 0°c down to −11°c, representing ~~the relation~~ between detector signal and anode power of the discharge power oscillator. An apparent double resonance occasionally observed was due to variation of the oscillator power which was not stabilized for this experiment.

3.3. *Electron Attachment in Iodine Discharge Afterglows*

In afterglow studies the electron concentration is changing with time and therefore the plasma resonance frequency is also time dependent. The galvanometer as detector of the probing signal has therefore to be replaced with a microwave receiver, amplifier and oscillograph display which is synchronized to the discharge pulse recurrence frequency. The plan of the experiment in this case consists in setting the probing signal frequency at a series of chosen values. For each of these probing frequencies a peak signal will be observed in the receiver at the time in the afterglow when the plasma resonance frequency reaches coincidence. A copy of a typical peak is shown in the inset in figure 5. Having recorded the relation between afterglow time and plasma resonance frequency the relation between time and electron density may be deduced. The temporal variation of electron density explored in this manner can be used to calculate the electron attachment rate provided conditions are chosen in which other processes of electron loss can be shown to be comparatively small.

Figure 5 is a graph of the logarithm of electron population against time in the afterglow for different pressures of iodine vapour. For the case in which electron attachment is the only process of electron loss we may write

$$\frac{dn}{dt} = -\beta n N_0$$

where β is the attachment coefficient and n and N_0 the concentrations of electrons and neutral iodine molecules respectively.

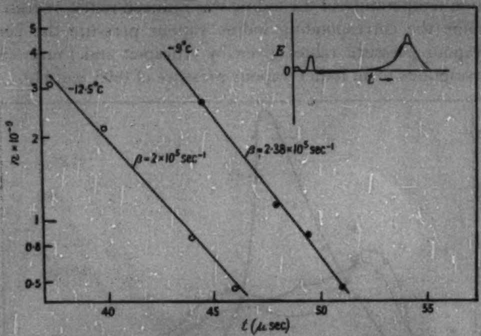


Figure 5. Decay of electron density in the afterglow.

The value of β was found to be of the order of 10^5 per second at pressure of the order of 10^{-2} mm Hg. An approximate calculation shows that the contribution of ambipolar diffusion to electron loss is not greater than 5% of this value and the diffusion contribution can therefore be neglected. Unfortunately, in the present case, the electron temperature was not known, but if it is assumed to be about 8000°K at 40 μsec after the discharge, then the cross section of attachment σ for iodine vapour will be $7 \pm 0{\cdot}5 \times 10^{-18}$ cm^2 this value is in agreement with those obtained by Buchdahl (1941) and by Healey (1938).

§ 4. CONCLUSION

There are certain restrictions to this r.f. probe technique: (i) to avoid serious collision broadening of the resonance curves the collisional frequency must be very much less than the frequency of the exploring wave; and (ii) measurements will be difficult whenever there exist v.h.f. plasma oscillations unless their amplitude is much smaller than the exploring wave amplitude or a sensitive receiver of sufficient selectivity is used.

However, the method can be used in a continuous high frequency electrodeless discharge where wire probe methods are invalid. It will also provide reliable measurements of electron concentration in the presence of both positive and negative ions where the interpretation of wire probe characteristics is very difficult.

REFERENCES

Buchdahl, R., 1941, *J. Chem. Phys.*, **9**, 146.
Gilespiet, L. J., and Fraser, L. H. D., 1936, *J. Amer. Chem. Soc.*, **58**, 2260.
Healey, R. H., 1938, *Phil. Mag.*, **26**, 940.

Transit Time Measurements of Charge Carriers in Amorphous Selenium Films

By W. E. SPEAR†
Physics Department, University of Leicester

MS. received 6th March 1957

Abstract. A technique similar to that used in crystal counter experiments is applied to the study of transit times of electrons and holes in thin evaporated films of amorphous Se carrying Au electrodes. Electron–hole pairs are generated near the top electrode by short electron pulses (or light flashes) of 10^{-7} sec duration. The pulses observed when electrons are drawn across the specimen show a perfectly linear rising edge. The measured transit times lead to consistent effective mobility values μ_-, between 4·7 and $5{\cdot}5 \times 10^{-3}\,\mathrm{cm^2\,sec^{-1}\,v^{-1}}$. For holes, $\mu_+ \simeq 0{\cdot}15\,\mathrm{cm^2\,sec^{-1}\,v^{-1}}$. The measurements are a direct confirmation of deductions made from previous experiments. An investigation of the temperature dependence of mobility shows that μ_- is predominantly controlled by thermal release from trapping states lying at a depth of 0·25 ev below the conduction band; for holes the corresponding energy is about 0·16 ev. The observed discharge effects indicate the presence of a range of deeper trapping states. Their effect on the field is normally negligible, but they are likely to act as recombination centres. Recombination processes play an important part in the region of excitation; they cannot be described by a bi-molecular model.

§ 1. INTRODUCTION

THE growing interest in crystal counters has, in recent years, led to a closer study of the charge carriers released by the incident elementary particles or γ-radiation. These experiments, which have been carried out mainly on diamond and crystals of the silver and thallium halides, have provided valuable information about carrier mobilities and their temperature dependence (Yamakawa 1951, Allemand and Rossel 1954, Brown 1955).

In the experiments described in this paper a technique similar to that used in the above crystal counter experiments has been applied to the study of transit times of electrons and holes in thin amorphous Se films. The difference lies in the fact that here the charge carriers are produced by short pulses of comparatively low energy electrons; thus the energy of the incident electrons, which determines the extent of the region of excitation within the specimen, is uniform and easily controllable.

The results show that in the case of amorphous Se it is possible to assign consistent effective mobility values to the charge carriers, although the measured transit times are predominantly controlled by trapping processes. The observations also confirm in a direct way some of the deductions made from previous experiments under steady electron bombardment (Spear 1956).

† On leave of absence at Purdue University, Lafayette, Indiana, U.S.A.

§ 2. Experimental Method

The experimental arrangement is illustrated in figure 1. The signal generator S produces a square wave with a frequency of 20 cycles per second which provides a trigger pulse for the time base of the oscilloscope and, with a delay of 0·5–2 μsec, triggers the discharge of a simple pulse-forming network P. The discharge pulse is applied between filament and control grid of the tetrode electron gun G, which is normally biased beyond its cut-off value. The bombarding pulses so obtained are of about 10^{-7} sec duration. The electrons are focused magnetically and enter the specimen through the evaporated gold electrode T. A steady electric field is maintained across the specimen by the potential v. The bottom electrode B is connected to the Y amplifier whose input resistance R_i is shunted effectively by the specimen resistance R_s and the capacity C_i, which is the sum of input and specimen capacities. C_i is adjusted to 350 pF in all the experiments; with $R_i = 1\,\mathrm{M}\Omega$ and $R_s \gg R_i$ this leads to a time constant of 350 μsec, which is much longer than the transit time τ to be measured. Thus between $t=0$ and $t=\tau$ the potential developed across R_i will be directly proportional to the total charge on C_i and the oscilloscope trace will represent the charge displacement due to the motion of carriers which were produced by the short excitation pulse at $t=0$. With maximum sensitivity and a bandwidth of 7·5 Mc/s, a pulse height of 1 cm corresponds to the displacement of $2{\cdot}2 \times 10^6$ electronic charges across the specimen. An increased sensitivity of $2{\cdot}2 \times 10^5$ electronic charges per cm at a reduced bandwidth of 2 Mc/s was used to determine N_p, the number of electrons in the excitation pulse.

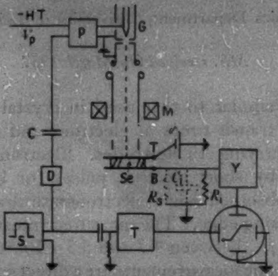


Figure 1. Experimental arrangement. S, square wave generator, T, triggered time base, D, delay, Y, Y-amplifier, T, B, specimen electrodes, C, high tension coupling condenser, P, pulse generator, G, electron gun, M, magnetic lens.

In some experiments electron–hole pairs were produced by photo-excitation. A spark of about 0·1 μsec duration was obtained by the discharge of a delay line across a spark gap in air (Beams 1947).

The experimental results with pulsed bombardment are not affected by the steady dark current which flows through the film in absence of electron bombardment (Spear 1956). In view of the comparatively large carrier densities near the top electrode, it is likely that secondary effects will develop. However, as has been shown previously, the growth of secondary currents is much too slow to contribute to the total charge displacement between $t=0$ and $t=\tau$.

§ 3. Experimental Results and Discussion

Altogether 15 specimens were investigated, ranging in thickness between 2 and 12 μ. The method of specimen preparation has been described previously. As compared with experiments under steady electron bombardment, the results with short excitation pulses are more reproducible and show far less variation between specimens. In addition, the lifetime of a specimen is practically unlimited.

Transit times were normally measured directly on the oscilloscope screen, although in some experiments traces were recorded photographically. Most measurements were made with $N_p = 10^5$ electrons per pulse, and with bombarding energies V_p between 7 and 9 kev. For these energies the average depth of penetration of the incident electrons p is about 2×10^{-5} cm. It may therefore be assumed that (i) all carriers which reach the bottom electrode have moved a distance d equal to the specimen thickness, and (ii) carriers of opposite sign contribute only a negligible amount to the total charge displacement. Owing to the large optical absorption coefficient of amorphous Se in the blue region of the spectrum, the same will apply to photo-excitation with blue light.

3.1. *Mobility at Room Temperature*

Figure 2 shows a number of typical pulses which were copied directly from photographs. The following observations apply to all $v > v_0$, where v_0 is of the order of 5 v:

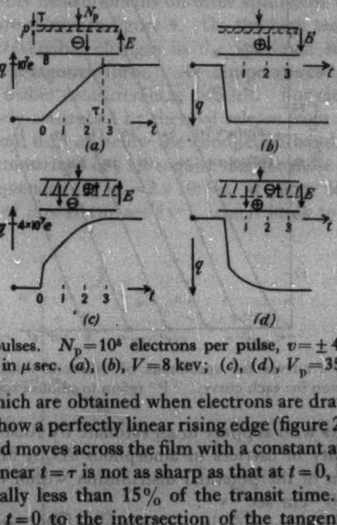


Figure 2. Observed pulses. $N_p = 10^6$ electrons per pulse, $v = \pm 40$ v, $d = 6{\cdot}95\,\mu$. Time scale in μsec. (*a*), (*b*), $V = 8$ kev; (*c*), (*d*), $V_p = 35$ kev.

(i) The pulses which are obtained when electrons are drawn out of the region of excitation always show a perfectly linear rising edge (figure 2 (*a*)). This suggests that the electron cloud moves across the film with a constant average drift velocity. The change of slope near $t = \tau$ is not as sharp as that at $t = 0$, but the extent of the curved part is normally less than 15% of the transit time. All measurements of τ are made from $t = 0$ to the intersection of the tangents to the linear and horizontal portion of the trace.

(ii) When the field across the film is reversed the height of the pulse remains unchanged. Thus for a given applied field the total number of holes drawn across the specimen per pulse is equal to that of the electrons. But figure 2 (*b*) shows that the transit time for holes is considerably shorter than that for electrons. This is a direct confirmation of the deductions made from previous measurements, which suggested that in amorphous Se volume generated currents are predominantly carried by positive holes.

(iii) For a given specimen it is found that $1/\tau_- \propto v - v_0$, where the subscript of τ refers to the sign of the carrier. The measured values of τ_- may be used to define an effective electron mobility μ_-. The term 'effective' is used to emphasize that μ_- is not a 'microscopic' mobility, but is based on measurements which include the total time during which an electron is trapped. On the assumption that $E = (v - v_0)/d$,

$$\frac{1}{\tau_-} = \mu_- \frac{v - v_0}{d^2}. \qquad \ldots\ldots(1)$$

In figure 3 six $(1/\tau_-, v/d^2)$ curves are drawn side by side for specimens between 2 and 12 μ thick. The curve marked P was obtained with photo-excitation. τ_- values were accurately reproducible up to the largest v used ($\sim$120 v). The values of μ_- are found to lie between $4{\cdot}7 \times 10^{-3}$ and $5{\cdot}5 \times 10^{-3}$ cm^2 sec^{-1} v^{-1}. Figure 3 shows conclusively that the concept of effective mobility describes a fundamental property of electrons in these films.

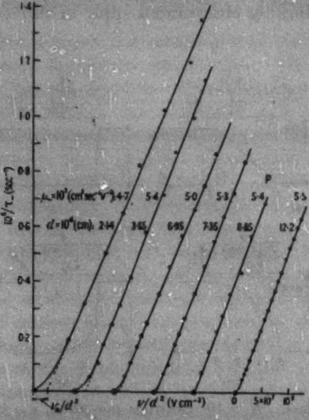


Figure 3. $(1/\tau_-, v/d^2)$ curves for specimens of different thickness *d*. The mobility values μ_- are given for each curve. 'P' refers to photo-excitation.

The above results lend support to an estimate by Ansbacher and Ehrenberg (1951) of the mobility of carriers in amorphous As_2S_3 films. It is concluded from a consideration of the potential distribution within the films that μ must be of the order of 10^{-3} cm^2 sec^{-1} v^{-1}.

At room temperature the mobility of holes cannot be determined by the above method because, except at small v, the transit time approaches the duration of the excitation pulse. From a number of experiments with small applied fields μ_+ is estimated to lie between 0·1 and 0·2 cm^2 sec^{-1} v^{-1}, so that $\mu_+/\mu_- \sim 40$. However, at lower temperatures, where τ_+ falls within the microsecond range, the pulses possess the same characteristics as the electron pulses.

(iv) It is interesting to extend the experiments to higher bombarding energies. With a larger value of p one would expect a contribution to the total charge displacement from both electrons and holes which should be easily distinguishable. This is shown by the traces in figures 2 (*c*) and 2 (*d*), photographed at $V_p = 35$ kev ($p \simeq 0{\cdot}6d$). Electron and hole contributions are unequal, because the integrated displacement of the type of carrier drawn towards B is in both cases larger than that of the carriers moving towards T. At bombarding energies somewhat larger than V_0 (the energy at which the incident electrons begin to pass through the film) electron and hole contributions become equal.

3.2. *The Temperature Dependence of Mobility*

The low mobility values, particularly for electrons, suggest that trapping effects play an important part; the total time during which a carrier is free to move may amount to only a small fraction of τ. Rose (1951) has given an expression relating μ and μ_0 the mobility expected in the absence of trapping states; μ_0 would then be determined by the mean free path and the average time between lattice collisions. If n_t is the electron trap density in a level of trapping states lying ϵ ev below the bottom of the conduction band, then in the neighbourhood of room temperature

$$\mu_- = \frac{10^{19}}{n_t} e^{-\epsilon/kT} \mu_0. \qquad \ldots\ldots(2)$$

The results of mobility measurements on three specimens in the temperature range 42°c to −30°c are shown in figure 4. On the high temperature side the range is limited by irreversible changes in the specimen such as the partial crystallization to the hexagonal form. At temperatures below −30°c the increasing value of v_0 makes measurements difficult. For electrons, the results indicate that μ_- is largely controlled by thermal release from a level of trapping states at a depth of about 0·25 ev below the conduction band. In addition, the experimental results summarized in § 3.1 require that for these states the trapping times $t_t \ll \tau$. From equation (2), $\mu_0 = 2 \times 10^{-17} n_t$ at room temperature. One might expect a value of $\mu_0 \sim 10\,\mathrm{cm}^2\,\mathrm{sec}^{-1}\,\mathrm{v}^{-1}$, which leads to a trap density of $5 \times 10^{17}\,\mathrm{cm}^{-3}$.

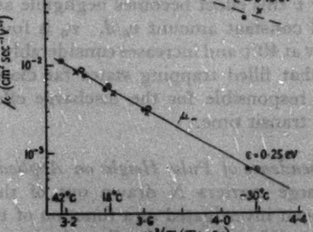


Figure 4. Temperature dependence of electron and hole mobilities.

As explained above, the results for holes shown in figure 4 can only be regarded as approximate. ϵ is of the order of 0·16 ev. If μ_0 and n_t are assumed to be the same for electrons and holes then the difference in the activation energies alone accounts for the observed ratio of μ_+/μ_-. It is possible, therefore, that the larger mobility of holes in Se is due to a smaller activation energy of the trapping states.

It is of interest to compare the above results with measurements on single crystals of Se grown from the vapour phase (Plessner 1951). Mobility values obtained from conductivity and Hall effect measurements showed that $\mu_+ \propto e^{-\epsilon/kT}$ with $\epsilon = 0{\cdot}13$ ev. At room temperature μ_+ was found to be 0·14 cm^2 sec^{-1} v^{-1}. In view of the difference in structure of the specimens the agreement of these values with those for amorphous Se is surprising. It suggests that the type of trapping centre controlling the hole mobility is similar in both the crystalline and amorphous forms.

3.3. *Polarization Effects*

A discharge effect has been observed which is of interest in connection with trapping processes. It is essentially similar to the polarization effects in counting crystals which may eventually lead to an inversion of the observed pulses (e.g. Wooldridge *et al.* 1947).

In this experiment normal pulses as shown in figure 2 (*a*) or 2 (*b*) are observed initially. The beam is now cut off and then v reduced to zero, so that both T and B are at earth potential. On switching the beam on again a pulse of the carrier of opposite polarity is observed which decreases rapidly in height and disappears within 1–2 seconds. The effect is observed with both signs of carriers and is most pronounced at low v (5–10 v) and values of N_p larger than those normally used. If, instead of switching the beam on immediately after removing v, an interval of 4–5 seconds is allowed, the discharge pulse is still about half its previous height.

These observations are clearly connected with a range of deep lying trapping states which satisfy the condition $t_t \gg \tau$. Unlike the trapping processes discussed in the previous section, these states will have little direct effect on the carrier mobilities. However, they will (i) modify the field within the specimen, and (ii) act as centres through which the recombination of carriers takes place. The effect of the trapped carriers is clearly shown in figure 3. At low applied potentials, $v < v_0$, the space charge reduces the effective field below the value v/d, and therefore leads to increased values of τ in this range. However, the results indicate that with larger v this effect becomes negligible and merely reduces the applied field by a small constant amount v_0/d. v_0 is found to be temperature dependent; it is about 1 v at 40°c and increases considerably at lower temperatures. It must be concluded that filled trapping states far closer to the steady state Fermi level than those responsible for the discharge effect, contribute to the value of v_0/d during the transit time.

3.4. *Dependence of Pulse Height on Applied Field*

The number of charge carriers N drawn out of the excited region per bombarding pulse has been investigated as a function of the applied field. For $v > v_0$, $N_+ = N_-$ always. (N, v) curves rise linearly from the origin and show a tendency towards saturation at the larger v. The number of carriers of either sign produced by each bombarding pulse may be estimated from the incident energy and the average energy required for the production of one electron–hole pair ($\simeq 20$ ev). The difference between this and N will give the number of carriers lost by recombination in the excited region. It is found that, with the comparatively large values of N_p used, more than half of the generated charge carriers recombine at the highest applied field. As the experiments with steady bombardment indicate, it is possible that saturation would be observed with lower values of N_p. In the range from $N_p = 10^5$ to 8×10^5 per pulse it is found that N_p is almost proportional to N; this indicates that a simple bi-molecular recombination model, which would lead to $N \propto N_p^{1/2}$ cannot be applied in this case.

In a recent paper Rose (1955) proposed a model in which recombination processes in insulators are essentially controlled by a range of trapping states (ground states), lying below the steady state Fermi level. The observation of the discharge pulses lends support to such a mechanism. Also it is apparent that if the density of deep lying ground states ($t_t \gg \tau$) in the excited region were much larger than the density of the generated carriers ($\simeq 4 \times 10^{13}$ cm^{-3} at $N_p = 10^5$ per pulse), one of the components of the recombination process would be practically independent of N_p. This should lead to $N \propto N_p$. It is likely that these conditions are fulfilled, but further investigation of density and distribution of trapping states in these films is essential before any more definite conclusions can be reached.

ACKNOWLEDGMENTS

The author wishes to thank Dr. A. N. Hunter for advice and the loan of an amplifier, and Dr. E. Matsukawa for helpful discussions.

REFERENCES

ALLEMAND, C., and ROSSEL, J., 1954, *Helv. Phys. Acta*, **27**, 519.
ANSBACHER, F., and EHRENBERG, W., 1951, *Proc. Phys. Soc.* A, **64**, 362.
BEAMS, J. W., 1947, *J. Opt. Soc. Amer.*, **37**, 868.
BROWN, F. C., 1955, *Phys. Rev.*, **97**, 355.
PLESSNER, K. W., 1951, *Proc. Phys. Soc.* B, **64**, 671.
ROSE, A., 1951, *R. C. A. Rev.*, **12**, 362 ; 1955, *Phys. Rev.*, **97**, 322.
SPEAR, W. E., 1956, *Proc. Phys. Soc.* B, **69**, 1139.
WOOLDRIDGE, D. E., AHEARN, A. J., and BURTON, J. A., 1947, *Phys. Rev.*, **71**, 913.
YAMAKAWA, K. A., 1951, *Phys. Rev.*, **82**, 522.

Contact Potential Difference Measurements by the Kelvin Method

By J. C. RIVIÈRE†
H. H. Wills Physical Laboratory, University of Bristol

Communicated by J. W. Mitchell ; *MS. received* 4*th October* 1956, *and in revised form* 13*th February* 1957

Abstract. Measurements of contact potential differences between clean metallic surfaces produced by condensation from the vapour, by the vibrating condenser modification of the Kelvin method, yield the following values for average work functions, on the basis of an assumed value of 4·565 ± 0·025 ev for the work function of tungsten: copper, 4·52 ± 0·04 ev; silver, 4·31 ± 0·03 ev; gold, 4·70 ± 0·03 ev; aluminium, 4·20 ± 0·05 ev; nickel, 4·74 ± 0·04 ev; molybdenum, 4·22 ± 0·05 ev; iron, 4·17 ± 0·03 ev. The lack of contamination and the reproducibility of results are attributed to the use of molybdenum gettering films, and to the outgassing procedures employed.

§ 1. INTRODUCTION

THE work described here formed the first stage in a programme whose ultimate aim was an investigation into the effects of certain reactive gases on the average work functions of clean metallic surfaces. Before these effects could be studied, it was necessary to determine accurately the work functions of the clean surfaces, since, with one or two exceptions, there is no agreement between the many values quoted in the literature for those quantities. The gas adsorption studies have not yet been brought to a satisfactory conclusion, but independent publication of results obtained with the clean surfaces is justified because of the reproducibility and accuracy of the measurements.

The fact that gases up to pressures of the order of a few millimetres are to be admitted at some stage in the experiment must influence the design of an experimental tube. Methods of measuring average (as opposed to photoelectric) work functions, that involve the use of an incandescent filament as a source of free electrons, cannot be used. In the first place, surfaces that consist of gas atoms or molecules adsorbed on a metal have a tendency to trap electrons, with the consequent creation of an abnormal surface dipole moment, and hence a spurious value of the work function. Secondly, the electron emission from incandescent tungsten filaments, such as are normally used, is unstable at relatively high pressures, particularly of oxygen. Even though a tungsten filament can be cleaned by flashing at 3000°K, films of gas start to form on the surface at the lower temperature to which the filament must be heated during measurements.

The alternative is to use the Kelvin method, which does not involve free electrons, and which has, moreover, certain other advantages. With the use of the Kelvin method, contact potential differences are measured directly and are not deduced from other experimental data, as is the case in the magnetron, diode, triode, and electron gun methods. Furthermore, the auxiliary electrodes that are a necessary but complicating adjunct of the latter methods are not required in the Kelvin method, so that it is probably capable of the greatest accuracy. For the same reason, there are fewer metal components in the system to be outgassed, and therefore less risk of contamination of the freshly prepared surfaces. The vibrating condenser modification of the Kelvin method was adopted in this work, and a detailed description of its application is given in later sections.

---

† Present address: Division of Industrial Chemistry, C.S.I.R.O., Fishermen's Bend, Melbourne.

The eight metals studied were copper, silver, gold, aluminium, nickel, iron molybdenum and tungsten. In a paper parallel to this, Mitchell and Mitchell (1951) described measurements made by the electron gun method due to Anderson (1935), of the contact potential differences between tungsten and each of the metals copper, silver and aluminium. The tungsten was in the form of a well-aged strip and the other metals were deposited separately from the vapour on to the strip. In order that results obtained by that method and by the Kelvin method might be compared, the first measurements made were of the contact potential differences between copper and silver, between silver and aluminium, and between aluminium and copper. They were followed by measurement of the contact potential difference between silver and gold. Gold was included primarily because it would later be required as a standard surface in atmospheres of oxygen (gold does not chemisorb oxygen at room temperature, as shown by Trapnell (1953)), but also because evaporated films of gold were being used in concurrent work of a different nature in this laboratory.

The four transition metals, nickel, iron, molybdenum and tungsten, were studied because they are the metals that have received most attention in catalytic theory. The behaviour of the work functions of these metals under all conditions of partial coverage from zero to unity of each of the four gases, nitrogen, hydrogen, carbon monoxide and oxygen, was to be investigated in the hope of obtaining information about the way in which each of these gases was adsorbed. From this information it was hoped to gain some insight into the mechanisms of certain of the simplest reactions that occur between gases at transition metal surfaces.

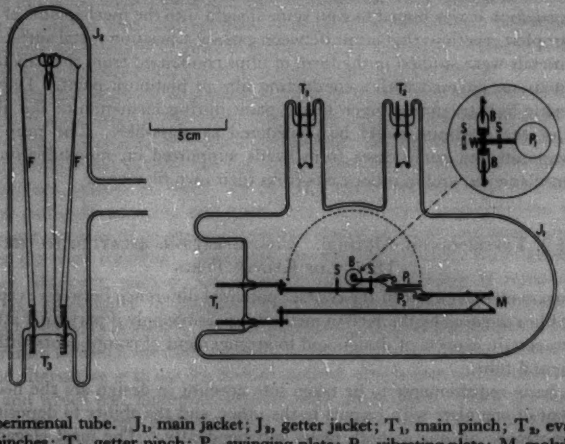
All metals were studied in the form of films condensed from the vapour on to glass substrates covered with a conducting film of platinum paint. Due to the very drastic heat treatment given to the paint during formation of the film, the surface of the platinum could be considered as glass-like. The pure metals were evaporated in some cases from beads supported on molybdenum or on tantalum filaments, and in other cases from their own filaments.

§ 2. EXPERIMENTAL METHOD. CONSIDERATIONS RELATING TO THE DESIGN OF KELVIN TUBES

Accurate measurement of the contact potential difference between evaporated metallic films *in vacuo* by the Kelvin method is possible only if particular attention is paid to certain aspects of design and to arrangement of components within the experimental tube.

The basic requirements to be taken into account in design are the necessary movement of one plate with respect to the other, and the ability to deposit metal from the vapour on one or both plates independently. The most elegant, and probably the most sensitive, modification of the Kelvin method is that due to Zisman (1932) and termed the 'vibrating condenser' method. Zisman's method of setting the vibrating plate in motion by the transmission of continuous vibrations through a metallic diaphragm in the wall of the tube could possibly be adapted to use under high vacuum conditions, but its adaptation would involve unnecessary complications in the construction of the tube. A method that involved the minimum of constructional difficulties was required for this work, since the research programme envisaged the construction of several score of identical tubes. The method adopted finally was to mount the vibrating plate on a rigid crossed-hinge suspension that allowed it to vibrate with a natural frequency of about 80 c/s when disturbed from its equilibrium position. Vibration was started by tapping the wall of the tube with a padded mallet.

In most early work that involved use of the Kelvin method (e.g. Glasoe 1931, Rose 1933, Farnsworth and Winch 1940) the motion of one plate away from the other for the purpose of contact potential difference measurement was also used to separate the plates for the deposition of metal on them. However, in the vibrating condenser modification, the plate whose vibrational motion is used for contact potential difference measurement is otherwise almost necessarily fixed in position, and an additional movement of the other plate away from it must be introduced. Mignolet (1950) arranged the suspension of the non-vibrating plate so that it could be pulled to one side by magnetic attraction, to allow evaporation on to the wall of the tube, which formed the other plate. On the other hand, Anderson (1952) mounted one of the plates on a wire framework, so that it could be slid up or down the length of the tube by gravity. Such a system was necessary in his design since the sliding plate had to assume three positions, corresponding to evaporation, to contact potential difference measurement by the electron beam method, and to contact potential difference measurement by the Kelvin method. Only two positions were required in this work, those for evaporation and for measurement by the Kelvin method, and the arrangement that was



The experimental tube. J_1, main jacket; J_2, getter jacket; T_1, main pinch; T_2, evaporator pinches; T_3, getter pinch; P_1, swinging plate; P_2, vibrating plate; M, molybdenum crossed-hinge; B, glass bearings; W, tungsten axle; S, U-shaped nickel supports; F, getter filaments.

developed is shown in the figure and its inset. P_1 is the non-vibrating plate, and it was mounted on the highly polished tungsten rod W, seated at each end in the glass bearings B. By suitable inclination of the tube P_1 could be swung into either of the two positions required of it. P_2 is the vibrating plate mounted on the crossed-hinge M of molybdenum strip. This arrangement could be assembled quickly and easily, and it employed few relatively massive metal parts in the tube.

Once the experimental design of a Kelvin condenser has been decided, careful attention must be paid to the problem of ' stray capacities '. Anderson (1952) has rightly pointed out that, in the theoretical treatment of the Kelvin method, complete isolation of the condenser is implicitly assumed, but that, in practice, the metal pair forms part of an array that includes the electrostatic shield and other possible components in the tube. No one applied potential can bring the field at the plates to zero, but the effect of extraneous fields is minimized by making the distances between either plate and its surroundings very large compared with the inter-plate spacing. The sharpness of the balance point will depend on the relative magnitudes of the change in capacity between the vibrating plate and its surroundings and the change in capacity of the Kelvin condenser. Anderson's conclusion is that sharp balance points can only be obtained for plate-to-shield spacings that are too large to be practicable in glass systems that have to be outgassed.

It is uncertain what dimensions Anderson has in mind for tubes that are too large to make outgassing practicable, for the tubes used in this work were of diameter 9 cm and length approximately 22 cm and no trouble was ever experienced in connection with their outgassing. Moreover, the sharpness of the balance point was sufficient to enable contact potential difference measurements to be made to ± 0·015 volt in the least sensitive tubes, and to ± 0·0005 volt in the most sensitive. The electrostatic shield surrounding the Kelvin condenser was a conducting film of platinum on the walls of the tube, and was equivalent to a shielding cylinder 9 cm in diameter and about 16 cm in length. It was held at earth potential throughout the measurements. Rough calculations show that, while the change in capacity of the Kelvin condenser during vibration was about 30 pF, the change in capacity between the vibrating plate and the shield was not greater than 0·002 pF. The relative magnitudes of these changes make it clear why the presence of the shield did not affect the sensitivity of the measurements, even when there was a difference of several volts between the potentials of the vibrating plate and of the shield. Nevertheless, reduction of the diameter of the shield by a significant amount (more than 2 cm, say) soon magnifies the effect of the contact potential difference between the vibrating plate and the shield to the point where a sharp balance is impossible. In this case Anderson's method of coating the back as well as the front surface of the vibrating plate, and also the area of tube wall near the vibrating plate, with the same metal, must be adopted.

§ 3. EXPERIMENTAL DETAILS

In this section will be described the details and construction of the experimental tube illustrated in the figure. It was constructed in two parts, the main tube containing the Kelvin plates and the metal evaporators, and a subsidiary tube containing the getter filaments.

The main jacket J_1 carried four side arms, two for the evaporators, one to carry the contact to the internal platinum shielding, and one as connection to the getter tube and pumping stem. Side arms for the evaporators of all metals except molybdenum and tungsten were of the size and shape shown in the figure, but due to the much higher temperatures reached during evaporation of the latter metals, tubing of diameter 3·8 cm was used for molybdenum, and the side arm was blown out into a sphere of diameter 6·1 cm for tungsten. The inner surface of the cylindrical part of J_1 was given three coats of Johnson, Matthey 'Bright Silver' platinum paint, electrical contact to which was provided by a spiral of tungsten wire. Three pinches were sealed into J_1, one of them, T_1, containing two tungsten leads on which were mounted the plate assemblies, and the other two, T_2, containing three tungsten leads each (only two shown in the figure) on which were mounted the evaporating filaments.

The plates were cut from Hysil plate glass and electrical contact to the metallic films to be deposited on them was provided by giving each plate three overall coats of platinum paint, each successive coat being burnt well into the glass. A length of platinum wire was wound tightly around a platinized spigot on the back of each plate and the free end of the wire spot-welded to the appropriate tungsten support. The crossed-hinge M was fabricated from two equal lengths of bright-annealed molybdenum strip, spot-welded to nickel rods on the tungsten supports while the latter were held parallel. Short lengths of glass tubing formed the bearings B in which W rotated, and they and their supporting tungsten rods were coated with 'Aquadag' to avoid accumulation of static charge *in vacuo*.

The copper, silver, gold, nickel and iron used in this work were supplied by Johnson, Matthey & Co. Ltd., in the form of wire of diameters 0·5, 0·5, 0·3, 0·2 and 0·5 mm respectively, and their purity was certified as better than 99·99%. Aluminium was supplied by the Aluminium Wire Co. in the form of 0·9 mm wire of 99·99% purity, molybdenum by Murex Ltd. in the form of 0·5 mm wire of purity stated to be better than 99·9%, and tungsten by British Thomson-Houston Co. Ltd. in the form of 0·3mm wire of purity stated to be better than 99·9%.

Beads for the evaporation of copper, silver, and gold were formed on hairpin filaments of 0·3 mm molybdenum wire by fusion in a stream of hydrogen. The short lengths of wire to be fused were first cleaned by etching for a few seconds in the appropriate concentrated acid, followed by washing several times in distilled water.

Nickel could not be formed into beads on hairpin filaments in the above way since it alloyed too rapidly with tungsten or molybdenum near its melting point. However, if the local proportion by weight of nickel to tungsten on a filament of the latter is kept below 30%, the nickel will evaporate at its melting point in preference to alloying (Olsen, Smith and Crittenden 1945). The construction of nickel evaporators in the form of coiled coils on 0·4 mm tungsten wire fulfilled this condition, and after construction the evaporators were cleaned by heating in the hydrogen stream to just below the melting point of nickel.

The most satisfactory filament for the evaporation of aluminium was found to be bright-annealed tantalum strip of thickness 0·1 mm. Lengths of aluminium wire were fused *in vacuo* on loops in the tantalum, immediately after being etched in cold concentrated hydrofluoric acid and washed in distilled water. In this way oxidation was reduced to a minimum.

Iron, molybdenum and tungsten were evaporated from their own filaments, which in all cases were made in the form of coils. After formation, all filaments were heated at appropriate temperatures in hydrogen.

The pinch T_2, sealed into the jacket J_2, carried four tungsten leads and a glass column, into the other end of which were sealed two tungsten spring hooks. Getter filaments F of 0·3 mm molybdenum wire were looped over the hooks and spot-welded to the pinch leads with the hooks under tension. The filaments were cleaned by prolonged heating at a high temperature in a hydrogen stream.

Apart from the tungsten–glass seals, for which Intasil glass was used, the tubes were fabricated from Hysil glass for the first thirty, and from Phoenix for the subsequent twenty-nine. All glassware was washed in nitric acid and distilled water after annealing. Components of nickel or molybdenum that were to be used in construction were hydrogen-furnaced at 900°c for 15 minutes. Any traces of grease on pinches, due to their handling during mounting, were removed by suspending the pinches in a reflux benzene degreaser. During assembly, a mixture of 9 parts nitrogen to 1 part hydrogen was streamed through the tubes to prevent oxidation, especially of the tungsten leads in the pinches.

Fifty-nine tubes were constructed in the course of this work, of which forty-three were of the design described here. Nineteen were discarded through accident or through faults in design or technique.

§ 4. Experimental Procedure

The outgassing procedure was similar in general to that of Myers (1952), and will be described only briefly here.

The tubes were exhausted with a single-stage mercury diffusion pump, backed by a two-stage rotary pump. There were liquid oxygen traps between the tube and the diffusion pump, and between the diffusion and backing pumps, and a phosphorus pentoxide trap between the second liquid oxygen trap and the backing pump. An ionization gauge was sealed to the system as close as possible to the tube; during the early part of this work a gauge of the conventional design was used, but for the major part one of the Bayard–Alpert type (Alpert 1953) was substituted, to considerable advantage.

Evacuation was carried on for about 75 hours, and during that time the tube was baked at 450°c for three spells totalling 50 hours. For the first five hours, the gauge and 'high-vacuum' trap were also baked at 450°c, after which the trap was immersed in liquid oxygen. Between the spells of baking, the evaporator and getter filaments were outgassed by prolonged heating at temperatures just below those at which evaporation of the metals was appreciable. The glass constriction was heated to the softening point, as many times as were necessary to produce no rise in pressure on further heating; three times were normally sufficient. The process of alternate baking and outgassing was continued until the ionization gauge was registering an ion current corresponding to about 10^{-8} mm Hg, and until further heating of the metal parts caused no rise in pressure. The time taken to reach this stage was invariably about 75 hours.

Immediately after the tube was sealed off from the pumping system, the getter filaments were heated separately to a temperature at which the molybdenum evaporated freely. Evaporation was allowed to continue until the molybdenum film on the walls of the jacket was black and opaque. The pressure after gettering, in those tubes that included gauges in their construction, could never be measured, since it was below the lower limits of sensitivity of both types of ionization gauge used.

The sealed-off tube was suspended rigidly inside a large earthed metal box. One of the leads from the Kelvin condenser was connected to a potentiometer capable of giving up to ± 1·9 volts at intervals of 0·0005 volt, and the other to an oscillograph via a pre-amplifier, and both were screened externally. After evaporation of the desired amounts of the two metals on to the plates, and movement of P_1 back into the measuring position, measurements were commenced by tapping J_1 with a soft mallet. The resultant signal that appeared on the oscillograph screen was balanced out by adjustment of the potentiometer setting, and at balance point, the potentiometer reading gave the numerical value of the contact potential difference but was of opposite sign. The readings were tabulated against time that had elapsed since the cessation of evaporation.

§ 5. RESULTS

The experimental results are summarized in the table, where they are arranged in two groups. Group I contains the contact potential differences between each of the pairs in the trio copper–silver–aluminium, while Group II is divided into two sub-groups, (*a*) consisting of the contact potential differences between silver and each of the metals nickel, tungsten, molybdenum and iron, and (*b*) between gold and each of the same metals. The sub-groups are connected by the contact potential difference between silver and gold, to maintain self-consistency.

(1)	(2)	(3)	(4)	(5)
	Cu–Ag	6	25	−0·205 ± 0·015
	Ag–Al	3	14	−0·110 ± 0·025
	Al–Cu	3	10	+0·315 ± 0·015
II (*a*)	Ni–Ag	6	30	−0·425 ± 0·020
	W–Ag	4	18	−0·252 ± 0·002
	Mo–Ag	6	27	+0·091 ± 0·035
	Fe–Ag	4	16	+0·144 ± 0·010
	Ag–Au	3	14	+0·386 ± 0·013
II (*b*)	Au–Ni	1	5	+0·039 ± 0·015
	Au–W	1	6	−0·134 ± 0·002
	Au–Mo	1	5	−0·477 ± 0·020
	Au–Fe	1	4	−0·530 ± 0·004

(1) Group; (2) metal pair; (3) no. of tubes; (4) no. of determinations; (5) contact potential difference (v).

The experimental limits of the contact potential difference in any tube were determined solely by the separation of the plates. This separation was fixed during construction, but it happened sometimes that the plate supports moved slightly during baking of the tube, causing increased separation and therefore reduction of the amplitude of the signal. In the worst case of this type, one of the molybdenum–silver tubes, the limits were ± 0·020 volt. In many individual determinations it was possible to balance the contact potential difference to ± 0·0005 volt.

§ 6. DISCUSSION

The most striking feature of these measurements was the constancy with time of the contact potential difference in any one experimental run. Owing to the nature of the method the interval between the end of evaporation and the taking of the first reading could not be reduced much below a minute, but from that time to over a week later the potentiometer setting did not have to be changed. In most cases the duration of an experimental run had to be restricted to a week, but in several cases where it could be allowed to go on for a longer period, there was no change in contact potential difference after four weeks, and in one case, none after seven weeks. This complete lack of contamination can be attributed partly to the preliminary outgassing procedures, but in much greater measure to the efficiency of the molybdenum getters. The contact potential difference measurements themselves furnish a very sensitive means of detecting the presence of adsorbable gases, for the work functions of either or both metal surfaces would change measurably in the presence of minute quantities of such gases. From the fact that there was no change in contact potential difference for at least a week, from the calculable time of formation of a monolayer at various pressures, and on the assumption that every gas molecule striking an uncontaminated metal surface sticks there, it is possible to show that the partial pressure of any adsorbable gas in the tubes must have been less than 10^{-12} mm Hg. The partial vapour pressures of inert gases such as helium, which Alpert and Buritz (1954) have proved diffuses through the glass walls of an evacuated vessel, may have been higher.

The function of ultra-high vacuum getters must be not just maintenance of the pressure in the tube at the pressure that obtained during seal-off, but reduction of the pressure as far as possible below the sealing-off pressure, for as long as possible. The latter task is clearly impossible if the getter itself releases gas during deposition, and a material must be used that can be outgassed thoroughly before its temperature is finally raised to the point of evaporation. Barium is unsatisfactory in this regard, for the temperature at which complete outgassing would be achieved is higher than the temperature at which evaporation is copious. There is always considerable evolution of gas during the actual ' gettering '. The high melting point metals tungsten, molybdenum and tantalum can be more readily outgassed than any others, but although Langmuir (1919) showed that molybdenum was able to act as a getter at relatively high pressures, nothing was known about the gettering properties of tungsten, molybdenum and tantalum films at low pressures until Dorling's (1953) work in this laboratory. Dorling's experiments showed that tungsten and molybdenum could be outgassed so completely that there was no evolution of gas during their subsequent evaporation, but that tantalum always gave up some gas when heated to the point of evaporation. In spherical bulbs in which tungsten and molybdenum films were deposited, the pressure fell from 10^{-6} mm Hg to below the lower limit of sensitivity of an ionization gauge, 10^{-8} mm Hg, and remained there for over a day. Heating of the bulbs to drive off gas caused the pressure to return to 10^{-6} mm Hg, but the pressure dropped again at once to the steady lower limit of the gauge when heating was stopped. Molybdenum films had the longest life under such heat treatment, and the lower temperature of evaporation from molybdenum filaments made them preferable to tungsten from the point of view of reliability. Dorling also found that films of molybdenum adsorbed about twice as many molecules of each of the gases O_2, CO, and N_2 per unit mass per unit geometric area of film, at room temperature, as did films of tungsten. As a consequence of Dorling's results molybdenum getters were introduced into ultra-high vacuum work in this laboratory by Schroeder in 1951, and their use was continued in this work. Their contribution to the technique of ultra-high vacuum physics is evident from the gratifying lack of contamination over long periods observed during the contact potential difference measurements.

Values for the average work functions of seven of the metals studied may be deduced if that of the eighth is known. The most reliable measurements of the work function of any metal by any method are those of well-aged polycrystalline tungsten by the thermionic method, and the accepted value is 4·545 ± 0·025 ev at 0°K (Ahearn 1933, Freitag and Kruger 1934, Wahlin and Whitney 1936, Reimann 1937). Due to its temperature dependence, $6{\cdot}5 \pm 0{\cdot}5 \times 10^{-5}$ ev deg^{-1} (Reimann 1937, Potter 1940, Mitchell 1950), the work function of tungsten becomes 4·565 ± 0·025 ev at 300°K. This value is likely to be nearer the work function of an evaporated tungsten film than the value obtained photoelectrically, since the thermionic method gives an average value for the work function, just as does a contact potential difference method, whereas photoelectric measurements refer to surface areas of minimum work function. Failing knowledge of the contact potential difference between a well-aged tungsten strip and an evaporated tungsten film, therefore, the work function of the tungsten films on glass used in this work will be put equal to the established value quoted above. Introduction of that value for the work function of tungsten into the contact potential difference results in the table gives at once the work functions of the other seven metals. The resultant values are compared in the list below with those of other workers; the figures quoted are those considered to be reliable from the points of view of correct application of method and, where it applies, of freedom from surface contamination. Photoelectric work functions have not been included in the list.

Copper	4·52 ± 0·04 ev.
	4·46 ev, evaporated film on glass (Anderson 1949).
	4·61 ev, evaporated film on tungsten (Mitchell and Mitchell 1951).
	4·50 ev, evaporated film on graphite (Jain and Krishnan 1953).
Silver	4·31 ± 0·03 ev.
	4·47 ev, evaporated film on glass (Anderson 1941).
	4·33 ev, evaporated film on tungsten (Mitchell and Mitchell 1951).
	4·31 ev, evaporated film on tantalum (Anderson 1952).
	4·31 ev, evaporated film on graphite (Jain and Krishnan 1953).
Gold	4·70 ± 0·03 ev.
	4·25 ev, evaporated film on graphite (Jain and Krishnan 1953).
Aluminium	4·20 ± 0·05 ev.
	4·25 ev, evaporated film on tungsten (Mitchell and Mitchell 1951).
Nickel	4·74 ± 0·04 ev.
	4·61 ev, strip (Wahlin 1942).
	4·96 ev, evaporated film on tungsten (Bosworth 1939).
	5·09 ev, strip (Cardwell 1949).
	4·50 ev, evaporated film on graphite (Jain and Krishnan 1952).
Molybdenum	4·22 ± 0·05 ev.
	4·15 ev, ribbon (Du Bridge and Roehr 1932).
	4·32 ev, wire (Ahearn 1933).
	4·33 ev, wire (Freitag and Kruger 1934).
	4·17 ev, wire (Wahlin and Reynolds 1935).
	4·20 ev, ribbon (Wright 1941).
Iron	4·17 ± 0·03 ev.
	4·21 ev, γ-strip > 1200°K (Wahlin 1942).
	4·48 ev, β-strip < 1200°K (Wahlin 1942).
	4·31 ev, γ-Fe film on graphite (Jain and Krishnan 1952).

Agreement between the values of the work functions obtained in this work and those measured by other workers is good in the cases of copper, silver, aluminium and molybdenum, fair in the cases of nickel and iron, and bad in the case of gold. Confirmation of the theory that contact potential difference measurements by the electron beam method give the same results as those by the Kelvin method is provided by the close agreement between these results and those of Mitchell and Mitchell (1951). Anderson has also provided confirmation from measurements of the contact potential difference between silver and barium by the two methods in the one tube.

Unless the nature of the substrate has a much larger influence on the work function of gold than it has on the work functions of copper and silver, it is difficult to account for the discrepancy between the result of this work and that of the effusion method of Jain and Krishnan. Additional accurate measurements are needed to resolve the difficulty.

The spread of previous work function values for nickel is greater than for any of the other metals, and may be due to the notorious difficulty in outgassing nickel completely. That outgassing may have been incomplete is also supported by the fact that the effusion method of Jain and Krishnan, which is not sensitive to surface contamination, gave the lowest value. On the other hand, the very high reproducibility of results in the present work seems to indicate that the nickel films were in fact clean, and the difference between the observed work functions may again be some function of the substrate surface.

In the case of iron, the difference between the observed values of the work function of a strip above and below the β–γ point is sufficiently greater than the experimental errors to be able to attribute characteristic work functions to each of the allotropic forms. The fact that the work function of the iron films studied by Jain and Krishnan (1952) is nearer that of γ-iron than of β-iron shows that some recrystallization must have taken place in the films, since they were evaporated initially on to cold graphite, and then heated to temperatures in the γ-range during the effusion measurements. However, in the present work, although the work function of the iron films is very close to that found for γ-iron by Wahlin, it is hard to see why the atomic arrangement in the films should resemble that in γ-iron, since the glass substrate was always cold relative to the β–γ point. This is another anomaly that must be explained by further work.

ACKNOWLEDGMENTS

The author would like to express his sincere thanks to Dr. J. W. Mitchell for his constant interest and guidance in this work, and his indebtedness to Mr. D. A. Jones, without whose skill in glassblowing this work could not have been attempted. He would also like to thank Mr. J. H. Burrow for his sound advice based on many years' experience in this field. The Executive of the Commonwealth Scientific and Industrial Research Organisation, Melbourne, are thanked for the award of an Overseas Research Studentship, which provided financial support for two years, and the University of Bristol for financial support in the third year.

REFERENCES

AHEARN, A. J., 1933, *Phys. Rev.*, **44**, 277.

ALPERT, D., 1953, *J. Appl. Phys.*, **24**, 860.

ALPERT, D., and BURITZ, R. S., 1954, *J. Appl. Phys.*, **25**, 202.

ANDERSON, P. A., 1935, *Phys. Rev.*, **47**, 958 ; 1941, *Ibid.*, **59**, 1034 ; 1949, *Ibid.*, **76**, 388 ; 1952, *Ibid.*, **88**, 655.

BOSWORTH, R. C. L., 1939, *Trans. Faraday Soc.*, **35**, 397.

CARDWELL, A. B., 1949, *Phys. Rev.*, **76**, 125.

DORLING, E. B., 1953, *Thesis*, University of Bristol.

DU BRIDGE, L. A., and ROEHR, W. W., 1932, *Phys. Rev.*, **42**, 52.

FARNSWORTH, H. E., and WINCH, R. P., 1940, *Phys. Rev.*, **58**, 812.

FREITAG, E., and KRUGER, F., 1934, *Ann. Phys., Lpz.*, **21**, 697.

GLASOE, G. N., 1931, *Phys. Rev.*, **38**, 1490.

JAIN, S. C., and KRISHNAN, K. S., 1952, *Proc. Roy. Soc.* A, **215**, 431 ; 1953, *Ibid.*, **217**, 451.

LANGMUIR, J., 1919, *J. Amer. Chem. Soc.*, **41**, 167.

MIGNOLET, J. C. P., 1950, *Disc. Faraday Soc.*, **8**, 326.

MITCHELL, E. W. J., 1950, *Thesis*, University of Bristol.

MITCHELL, E. W. J., and MITCHELL, J. W., 1951, *Proc. Roy. Soc.* A, **210**, 70.

MYERS, H. P., 1952, *Proc. Roy. Soc.* A, **215**, 329.

OLSEN, L. O., SMITH, C. S., and CRITTENDEN, E. C., 1945, *J. Appl. Phys.*, **16**, 425.

POTTER, J. G., 1940, *Phys. Rev.*, **58**, 623.

REIMANN, A. L., 1937, *Proc. Roy. Soc.* A, **163**, 499.

ROSE, B. A., 1933, *Phys. Rev.*, **44**, 585.

TRAPNELL, B. M. W., 1953, *Proc. Roy. Soc.* A, **218**, 566.

WAHLIN, H. B., 1942, *Phys. Rev.*, **61**, 509.

WAHLIN, H. B., and REYNOLDS, J. A., 1935, *Phys. Rev.*, **48**, 751.

WAHLIN, H. B., and WHITNEY, L. V., 1936, *Phys. Rev.*, **50**, 735.

WRIGHT, R. W., 1941, *Phys. Rev.*, **60**, 465.

ZISMAN, W. A., 1932, *Rev. Sci. Instrum.*, **3**, 367.

The Neon Flash Tube as a Detector of Ionizing Particles

By M. GARDENER, S. KISDNASAMY, E. RÖSSLE† and A. W. WOLFENDALE
Department of Physics, The Durham Colleges in the University of Durham

Communicated by G. D. Rochester; MS. received 14th February 1957, and in final form 2nd April 1957

Abstract. A new type of particle detector developed recently by Conversi and his co-workers consists of a glass tube filled with neon which, after traversal by an ionizing particle, can be made to glow by the application of a strong electric field. In this paper details are given of a systematic study of the 'flash tubes' with a view to the production of a particle detector of high efficiency and low cost. It has been found that the requirements for high efficiency are not rigorous: relatively impure gas can be used and time delays of several microseconds between the traversal of the particle and the application of the field can be tolerated.

The limitations imposed by the presence of spurious flashes, i.e. flashes not associated with ionizing particles, are discussed.

§ 1. Introduction

The reason for the detailed study of the neon flash tubes was the need to construct a suitable particle detector for a high energy cosmic ray spectrograph. In a spectrograph the trajectory of a particle traversing a magnetic field is determined by finding the position of the trajectory at several levels distributed spatially along the path of the particle. In order to define the position accurately it is essential that the detector at each level should have a large area, and a high spatial and time resolution. Moreover, it is necessary to have a device which is relatively simple and of low cost. Previous spectrographs have met two of these requirements. Thus instruments using Geiger counters (Hyams *et al.* 1950) can have large collecting areas with good time resolution, and those with cloud chambers as detectors (Holmes *et al.* 1956) have good spatial resolution, moderately large areas, but poor time resolution. No spectrograph has so far met all requirements. Some form of parallel plate spark counter (e.g. Pidd and Madansky 1949, Robinson 1953) might seem to offer a solution. In these counters the time resolution is less than 10^{-8} sec and the position of the visible spark reproduces that of the ionizing particle to a fraction of a millimetre. Unfortunately, however, it has not been possible to construct counters of large area and long life.

The use of electrodeless gas-filled tubes as particle detectors was first described by Conversi and Gozzini (1955). The characteristics of the tubes have been described in more detail in later publications by Conversi and his collaborators (Conversi *et al.* 1955, Barsanti *et al.* 1956). Basically, the counter consists of a closed glass tube containing a gas or mixture of gases, of which neon is the main constituent. The tube is placed with others in a layer between parallel conducting plates, and when a selected ionizing particle traverses the layer of tubes a high voltage pulse is applied across the plates. The electrons produced in the gas of the tube through which the particle has passed are accelerated by the pulsed electric field, and the ensuing electron avalanche causes a luminous discharge which fills the whole tube.

---

† British Council Fellow, on leave from Hochspannungslaboratorium, Hechingen.

An array of flash tubes, each of small diameter, will satisfy the requirements set out above, and their use in a spectrograph should result in an instrument having high particle flux, high maximum detectable momentum, and low rate of spurious events. In this paper a description is given of an investigation of some of the parameters which affect tube characteristics with a view to the design of an efficient flash tube which could be mass-produced cheaply.

The tubes made by Conversi were of thin walled soda glass and the gas fillings were successively pure neon, pure neon and a small amount of argon, and commercial neon, the maximum pressure being about 45 cm Hg. An investigation of the properties of the tubes showed that the efficiency increased with the strength of the pulsed electric field, gradually levelling off to a plateau. As the time delay between the passage of the particle and the application of the pulse was increased, the efficiency fell off, the decrease being very rapid for commercial neon if the delay was greater than about 0·3 μsec. For given pulse characteristics it was found that the addition of argon greatly lowered the threshold value of the field required to reach the plateau, without lowering the value of the efficiency on the plateau. As the pressure of the gas filling was increased, the efficiency increased to a maximum value at about 35 cm Hg after which there was a decrease. It was found necessary to evacuate the tubes to a pressure of less than 10^{-5} mm Hg before filling.

The particular object of our experiments was to see if the requirements for the construction and operation of the tubes need, in fact, be as stringent as those found by Conversi.

§ 2. THE EXPERIMENTAL PROCEDURE

2.1. *The Filling System*

A system was constructed in which an ultimate air pressure in the region 10^{-5} mm Hg could be achieved, where the tubes could be outgassed by baking, and where small known quantities of other gases could be mixed with the neon. Before filling, the tubes were baked for about 30 minutes at about 300°c. After filling, the tubes were sealed off and wrapped in black paper, so that each tube was shielded from the light from its neighbours.

2.2. *The Characteristics under Investigation*

The basic requirements of the tubes are that they should have a high efficiency for flashing when a particle passes through the array within the resolving time of the apparatus—a few microseconds—and should give a small number of spurious flashes, i.e. flashes not due to the passage of a particle. Two efficiencies may be defined: the layer efficiency, defined as the ratio of the number of single flashes observed in a layer to the number of single particles passing through it, and the absolute efficiency where only particles passing through the gas in the tubes are considered.

These efficiencies are a function of a large number of parameters, which can be grouped as follows:

(i) The dimensions of the glass tubes and the type of glass used.

(ii) The pressure of the main constituent of the gas filling (neon), the pressures of contaminants unavoidably present, and the pressures of gases added to the neon intentionally.

(iii) The properties of the high voltage pulse: the peak height, the rise time, pulse length, and the time delay between the passage of the ionizing particle and the commencement of the pulse.

Tubes have been constructed of Pyrex and soda glass. The former is easier to work and tubes of precise diameter can be obtained commercially (so-called 'precision-bore' tubing). Soda glass on the other hand, can be obtained with considerably smaller wall thickness, and so higher layer efficiencies can be achieved.

For high spatial resolution tubes of small diameter are required. For small diameter tubes, however, the wall thickness is roughly constant, independent of diameter, and so the ratio of gas to glass volume decreases as the diameter is reduced and the maximum layer efficiency falls off.

Another disadvantage with tubes of small diameter is that a high pressure of gas is required to give a reasonable number of ion-pairs from the ionization of the incident particle. In the investigations reported here the tubes were of the following dimensions:

Glass	External diameter	Internal diameter	Length
Pyrex	7·85 mm	5·75 mm	~15 cm
Soda	7·29 mm	5·58 mm	~15 cm

The primary gas used in the tubes was either spectroscopic or commercial neon. Spectroscopic neon (abbreviated Ne(S)) contains 99·8% neon and traces of other elements, whereas commercial neon (abbreviated Ne(C)) contains 98% neon, 2% helium, and small quantities of other elements.

2.3. *The Experimental Arrangement*

The efficiencies of tubes were determined using the apparatus shown in figure 1. Four-fold coincidences were selected, most of which were due to the passage of single particles through the tube assembly. The high voltage pulse was produced in a manner similar to that described by Conversi and Gozzini,

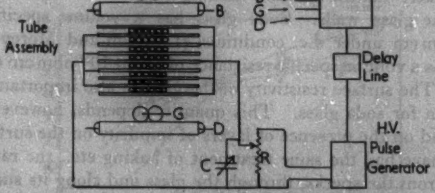


Figure 1. The experimental arrangement.

viz. an open delay line was discharged through a hydrogen-filled thyratron and the primary of a pulse transformer. The delay T_D between the passage of the particle and the commencement of the pulse was controlled by the delay line between the coincidence circuit and the high-voltage pulse generator. The height of the pulse on the plates was varied by adjusting the tapping on the matching resistance R, and the rise time T_R by varying C. The pulse length, τ, was varied by altering the components of the pulse forming network. (A typical pulse is shown in figure 3; τ is the pulse width at 75% of the peak height and T_R is the rise time to the 75% level.)

Layers of six similar tubes were placed in the plate assembly so that successive layers had different gas fillings. Photographs of the flashes were taken and the layer efficiency determined for each layer.

The choice of the values of the variables T_D and T_R to be used in the investigation was made bearing in mind the values likely to be attainable in a practical application of the technique. For investigations in which a number of layers of Geiger counters several metres apart are connected in coincidence to control the operation of the flash tubes, delays of several microseconds between the passage of the particle and the generation of the coincidence pulse may occur. Consequently, rather a long delay, $T_D = 3{\cdot}7\,\mu$sec, was used in most of the investigations. Similarly, a relatively long value of T_R is often dictated by the necessity of applying the high voltage pulse to a large array of tubes having considerable electrostatic capacity.

§ 3. THE EXPERIMENTAL RESULTS

3.1. *The Production of Spurious Flashes*

Considerable difficulty was experienced in the operation of tubes constructed of Pyrex glass for both Ne(C) and Ne(S). It was found that at moderate field strengths of about 6–8 kv cm^{-1} the tubes flashed frequently when subjected to random pulses i.e. pulses not associated with ionizing particles. It was found later that the region of field strength for which the efficiency is high and the flashes sufficiently intense for good photography is 6–10 kv cm^{-1}. Pyrex tubes are, therefore, unsatisfactory.

The spurious flashing of soda glass tubes was also studied. About 10% of the tubes flashed spuriously below 12 kv cm^{-1}; these were rejected. Of the remainder only about 10% flashed spuriously in the region 12–15 kv cm^{-1}. The soda glass tubes are therefore satisfactory.

It seems likely that the spurious flashing is due to the discharge from charged specks on the glass walls. Soda glass has a volume specific resistance of about 10^{13} ohm cm under d.c. conditions (Littleton and Morey 1933) whereas Pyrex glass has a volume specific resistance of about 10^{15} ohm cm (manufacturers' data sheet). The surface resistivity of the glass is also important, being greater for Pyrex than for soda glass. This quantity depends, however, on the grade of the glass and on the presence of layers of impurity on the surface. For glass tubes which have had the same treatment of baking etc., the rate of leakage of charge away from the specks, through the glass and along its surface, would be expected to be higher for soda glass, and consequently the threshold for spurious breakdown higher: this is in accordance with experiment.

Another source of spurious breakdown is the ambient radiation: the tubes are photosensitive. It has been found that a relatively low intensity of light, particularly in the blue end of the spectrum, produces an enhanced rate of spurious flashes. It seems probable that the source of the initiating electrons for the spurious flashes is photoelectric emission from the surface of the glass; a photon of blue light, $h\nu \sim 3$ev, has higher energy than the photoelectric work function of several constituents of glass (Na, Ca and K).

3.2. *The Reproducibility of the Tube Characteristics*

Before considering the way in which the efficiency of the tubes depends on the nature of the gas filling it is necessary to study the reproducibility of the fillings. Figure 2 shows the variation of layer efficiency with field strength for two sets of tubes filled under very similar conditions. It will be noticed that, although the shapes of the curves are similar, one is higher by nearly 9%. This lack of complete reproducibility must be borne in mind when detailed comparisons of characteristics are made.

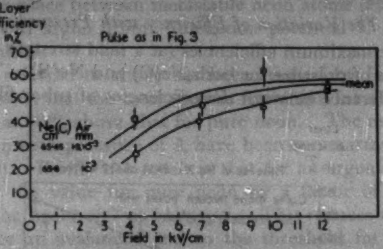


Figure 2. Characteristics of tubes filled under similar conditions.

3.3 *Variation of Efficiency with Air Pressure for Tubes Filled with* Ne(C)

Preliminary experiments showed that Ne(C) could be used as a tube filling and that the efficiency increased with pressure. Tubes were, therefore, filled with Ne(C) at the maximum pressure investigated, about 65 cm Hg. The variation of efficiency with field, with air pressure as parameter, is shown in figure 3, together with the characteristics of the high voltage pulse. The reason for this measurement was to see to what pressure the tubes must be evacuated before being filled. The effect of adding air is seen to be a reduction in the

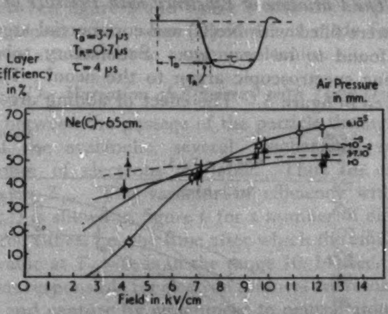


Figure 3. Variation of efficiency with residual air pressure.

maximum efficiency and a reduction in the threshold for high efficiency of flashing. For as much as 1 mm of air, the reduction in maximum efficiency, at 12 kv cm^{-1}, is only about 20%.

The reason for part at least of this reduction in efficiency is that the presence of oxygen increases the chance of there being no electron in the sensitive volume by the time the pulse is applied, as electrons are lost by attachment to the oxygen molecules. A quantitative estimate of the effect is given later in § 3.6.

The interpretation of the effect of air on the threshold for high efficiency of flashing is discussed in § 3.5.

3.4. *The Variation of Efficiency with Pressure of Neon*

In figure 4 the maximum efficiencies (at a field strength of 12·2 kv cm^{-1}) are given as a function of pressure for both Ne(C) and Ne(S). There appears to be no appreciable difference between the efficiencies of tubes filled with Ne(S) and Ne(C) in the region of 45 cm Hg, all the points fitting on one smooth curve. It can be seen that there is a marked increase of efficiency with pressure of gas. This occurs because a larger number of ion pairs is produced and so the loss of electrons by various mechanisms becomes less important (see § 3.6).

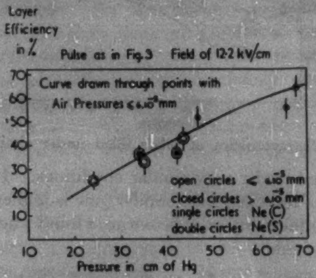


Figure 4. Variation of efficiency with pressure of neon.

3.5. *The Variation of Efficiency with Pressure of Argon*

Several tubes were filled with Ne(C) and commercial argon but the resultant efficiencies were found to be very low. Satisfactory tubes, were, however, produced by adding spectroscopic argon to the neon. The characteristics of these tubes are given in figure 5. The effect of adding argon is similar to that

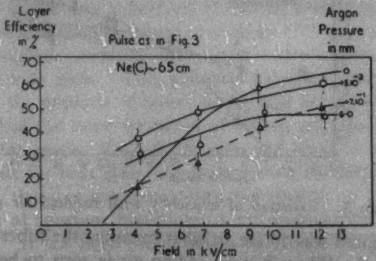


Figure 5. Variation of efficiency with pressure of argon.

of adding air: the maximum efficiency is reduced but there is an extension of the region over which the efficiency is high. At fairly low field strengths—below about 4 kv cm^{-1}—there is an optimum argon pressure at which the efficiency is a maximum. This pressure is in the region of $1{\cdot}5 \times 10^{-2}$ mm Hg, i.e. a fraction of $2{\cdot}3 \times 10^{-5}$%.

The increase in efficiency caused by the addition of small quantities of argon is due to two effects: the increase in the total specific ionization, and the increase in the ionization coefficient η — the probability of ionization by an electron per volt potential difference. Both effects arise from the fact that collisions of the second kind can take place between metastable neon atoms (excitation potentials 16·53 ev, 16·62 ev) and argon atoms (ionization potential 15·8 ev) whereby the argon atoms are ionized.

The first effect referred to above has been studied by Jesse and Sadauskis (1952) who have shown that the total specific ionization for a mixture of Ne and A may be as much as 40% above that for pure neon. The experimental results on the variation of η with pressure of A have been summarized by Druyvesteyn and Penning (1940). Briefly the result is that for an argon content of 10^{-2}%, η is increased over the value for pure neon by a factor of the order of 3 at 12 kv cm^{-1} and of the order of 7 at 6 kv cm^{-1}. This reduces the electron energy required to produce an avalanche and so the threshold for high efficiency of flashing is correspondingly reduced.

The pressure of argon for which the efficiency is a maximum depends on the magnitude of the applied field. At 4 kv cm^{-1} the optimum pressure is of the order of 2×10^{-3}%. Figure 5 shows that this optimum tends to zero as the field is increased. Kruithof and Penning (1937), however, showed that the pressure of argon giving the maximum value of η varies from 0·6% at 5 kv cm^{-1} to 1·2% at 13 kv cm^{-1}. This trend with field strength is in the opposite direction to that found here and the reason for the reduction in efficiency at high field strengths with increasing pressure of argon is not known. Further experimental study of this phenomenon is necessary.

The reason for the increase in efficiency at low voltage in the case of the tubes filled with neon and air can now be understood. Air contains about 0·9% argon and comparison between figure 5 and figure 3 with the air pressures multiplied by 10^{-2} yields agreement when some allowance is made for the reduction in efficiency by electron attachment to oxygen molecules.

3.6. *The Variation of Efficiency with Time Delay*

In the interval between the passage of the particle through the tube and the commencement of the avalanche, several mechanisms are responsible for reducing the number of electrons available. Thus the efficiency falls with increasing time delay T_D. The variation of efficiency with T_D at high field strength (11 kv cm^{-1}) is shown in figure 6 for a number of different tube fillings. The 'half-life' of the tubes, i.e. the time after which the efficiency falls to half of its (extrapolated) value at $T_D = 0$, is in the range 10–14 μsec.

The mechanisms responsible for electron removal are recombination, diffusion to the glass walls, and capture by attachment to neutral atoms to form negative ions. If the number of positive and negative ions per cm^3 are n_+ and n_- respectively, the number of recombinations per second is $\alpha_r n_+ n_-$, where α_r is the volume recombination coefficient. For neon, $\alpha_r \sim 2 \times 10^{-7}$ cm^3 ion^{-1} sec^{-1}

(Biondi and Brown 1949) and n_+ and n_- for a track at near minimum ionization in neon at 65 cm Hg pressure are each of the order of 20 ion cm^{-3}. It is apparent that the contribution to electron removal from volume recombination is negligible. Preferential recombination, i.e. the recombination of the original members of the ion pair, might be more important but even this effect should be small as the free electrons produced in the ionization process have appreciable energy and soon escape from the neighbourhood of their positive ions.

The diffusion of ions and electrons has been considered by many workers. If N_0 ions are concentrated at the origin at $t=0$, then after time t, the number between x and $x+dx$ is

$$N_x dx = \frac{N_0}{(4\pi Dt)^{1/2}} \exp\left(\frac{-x^2}{4Dt}\right) dx, \qquad \ldots\ldots(1)$$

where D is the diffusion coefficient.

In the case of electrons produced in a 'flash tube' by an ionizing particle traversing a diameter, if N_0 = the number of electrons produced, N_t = the number remaining within the tube, i.e. not having reached the wall, after a time t, and R = internal radius of the tube, then

$$N_t = N_0\left(1 - \exp - \frac{R^2}{4Dt}\right) f\left(\frac{R^2}{4Dt}\right) \qquad \ldots\ldots(2)$$

where $f(R^2/4Dt)$ is a geometrical function, determined graphically. ($f(R^2/4Dt)=1$ if it is assumed that the electrons are all produced at the centre of the tube.)

The varying path lengths of primary ionizing particles passing through the tube can be allowed for to a first approximation by using for N_0 the number of electrons produced on the mean path length $\bar{L}$. This number must be reduced still further, however, because of electrons produced so near to the wall that on acceleration they are unable to produce an avalanche (following Conversi *et al.* 1955). The internal efficiency of a tube for recording ionizing particles is then, approximately,

$$\eta_t = \eta_0\left[1 - \exp\left\{-N_t\left(1 - \frac{\delta}{\bar{L}}\right)\right\}\right] \qquad \ldots\ldots(3)$$

where η_0, the efficiency at zero time delay, is given by

$$\eta_0 = 1 - \exp\left[-n\left(1 - \frac{\delta}{\bar{L}}\right)\right] \qquad \ldots\ldots(4)$$

and is the probability of at least one primary ion-pair being produced in the sensitive volume. n is the average number of *primary* ion-pairs produced along the mean path length. δ is the correction for the insensitive region and is found by extrapolating the (η, T_D) curve back to $T_D=0$, and then using equation (4). The value of δ found in this way is 1·4 mm.

The expected efficiency has been evaluated for the tubes filled with Ne(C) at 66·5 cm Hg and 6×10^{-5} mm Hg of air. The total ionization has been derived from the work of Jesse and Sadauskis (1952) on the energy required to liberate one ion-pair, and the data of Eyeions *et al.* (1955) giving the most probable total energy liberated in the form of ionization in neon. At a mean momentum of 1 GeV/**c** the most probable total specific ionization is 21 ion-pairs cm^{-1}.

The theoretical curve for tubes filled with 66·5 cm of Ne(C) and having a residual air pressure of 6×10^{-5} mm Hg, is also shown in figure 6. It will be seen that it agrees fairly well with the experimental curve for time delays up to about 10 μsec. The divergence for longer time delays can be attributed to an approximation in the diffusion theory.

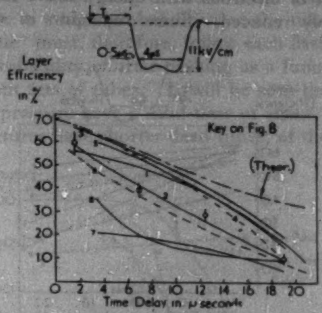


Figure 6. Variation of efficiency with time delay T_D.

Where gases having high electron attachment probabilities are present, removal of electrons by attachment is important. The most common impurity, introduced intentionally in several sets of tubes, is oxygen. The electron attachment coefficient h at thermal energies is about $2{\cdot}5 \times 10^{-4}$. As the electron energy is increased, h falls to a minimum at about 1·3 ev and rises to a maximum of $3{\cdot}5 \times 10^{-4}$ at 2·0 ev (Healey and Kirkpatrick, quoted by Healey and Reed 1941). Thus attachment is important in two regions: during the thermal motion before the pulse is applied, and at the stage of electron acceleration when the electron energy is in the region of 2 ev. An approximate calculation shows that after a delay of 10 μsec about 10% of the electrons will be attached to atoms in tubes containing 1 mm air and 67 cm Ne(C). A qualitative explanation of figure 6 is therefore possible. The fall of efficiency with time delay is more rapid for tubes containing air, and for tubes filled at the lower gas pressures. It is interesting to note that with a time delay of 2 μsec the efficiency of the tubes containing 1 mm Hg of air is virtually as high as that for those with an air pressure of less than 10^{-5} mm Hg.

3.7. *The Variation of Efficiency with Rise Time*

The electron drift starts with the commencement of the high voltage pulse. If the voltage increases slowly with time the electrons will drift to the wall before their velocity is sufficiently high to produce further electrons by ionization, and an avalanche will not be initiated. The efficiency of the tubes, would, therefore, be expected to fall with increasing T_R. Figure 7 shows this variation.

An approximate quantitative treatment of the variation is as follows. If w is the drift velocity produced by a field E then over a wide range of field strengths $w = kE$ where k is a constant, $\sim 2 \times 10^6$ for E in kv cm^{-1} and w in cm sec^{-1}. The variation of mean excitation energy ϵ with E, for pure neon. is such that ϵ increases up to a plateau of about 12 ev above $E = 3$ kv cm^{-1} (Rossi and Staub 1949). If the leading edge of the pulse is given by $E = \alpha t$ with α constant then the time for an electron to drift a distance s is $t = (2s/k\alpha)^{1/2}$. With $T_R = 2$ μsec, say, and $E_{\max} = 11$ kv cm^{-1} (figure 7), $\alpha = 4{\cdot}1 \times 10^6$, and, for s equal to the radius of the tubes, $\sim$0·28 cm, the value of t is 0·26 μsec. In this time the applied field rises to about 1·1 kv cm^{-1} for which $\epsilon \sim 9$ ev; this value is significantly below the maximum value and the number of electrons with energy above the ionization potential, 21·6 ev, is considerably reduced. Some reduction in efficiency is therefore expected for values of T_R of the order of 2 μsec. The greater reduction in the case of the tubes containing 1 mm air is presumably due to the increased probability of electron attachment during the 'rise-time' of the pulse, since the attachment coefficient has a resonance for electrons in the region of 2 ev (see § 3.6).

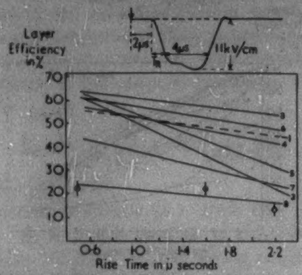


Figure 7. Variation of efficiency with rise time T_R. (Key in figure 8.)

3.8. *The Variation of Efficiency with Pulse Length*

At 11 kv cm^{-1} the drift velocity is of the order of 10^7 cm sec^{-1} and the time required for an electron to traverse the tube is of the order of 0·05 μsec. Consequently, no appreciable variation would be expected for pulse lengths greatly in excess of this value.

The efficiency as a function of field strength has been studied for pulse lengths of 2 and 4 μsec. The results for the 4 μsec pulse have been considered previously. Those for the 2 μsec pulse are given in figure 8. Comparison with figures 3 and 5 shows that there is no significant reduction in efficiency for the tubes filled at higher pressures. At lower pressures, however, there is an appreciable reduction in efficiency. The reason for this reduction is not known.

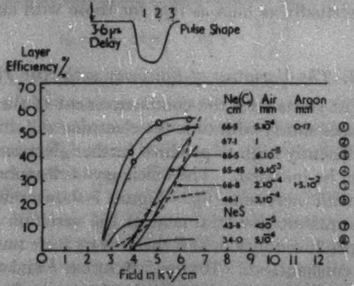


Figure 8. Variation of efficiency with field.

3.9. '*After-flashing*'

If a flash occurs in a particular tube due to the passage through it of an ionizing particle and then, after a short delay of the order of 1 sec, another high voltage pulse is applied, there is a high probability that the tube will flash again. An appreciable 'dead-time' must, therefore, follow each flash.

In figure 9 the probability of after-flashing as a function of time interval is given for four different sets of tubes. It will be seen that the tubes filled with 'pure' neon at high pressure have a dead-time of about 10 seconds whereas the tubes containing mixtures have shorter dead-times, of the order of 1 second.

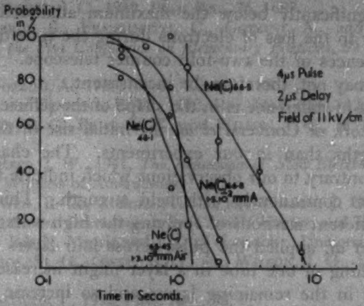


Figure 9. Probability of 'after-flashing' as a function of time.

The reason for the phenomenon of 'after-flashing' is that considerable time is taken for the very large number of electrons, produced in the avalanche, to be effectively removed from the gas in the tube. The probability of a few electrons remaining after a time of the order of 1 second is thus high. The mechanisms of removal are those given in § 3.6; immediately after the avalanche recombination is most important; when most of the electrons have been removed the other two mechanisms are more important and the conditions considered in § 3.6 are reproduced.

§ 4. DISCUSSION

The results reported here can now be compared with those of Conversi and his collaborators. The overlap between the two investigations is not very great, however, as the tubes studied by Conversi *et al.* were at low pressure (35 cm Hg) and operated with short time delays ($<1\,\mu$sec), whereas our tubes covered a pressure range of 35 cm–65 cm Hg, and the time delays were in general much longer ($\sim 3{\cdot}7\,\mu$sec).

In spite of the rather different conditions of the two experiments, the two sets of results show agreement in a number of respects. The maximum efficiency with 34 cm Ne(S) is consistent with that found by Conversi *et al.* for the same time delay, after allowance has been made for the different wall thickness of the tubes. In both cases the effect of adding argon is to cause a reduction in the threshold field strength for high efficiency of flashing. The maximum efficiencies with pulses of lengths $2\,\mu$sec and $4\,\mu$sec in general showed no significant difference for either our measurements or those of Conversi *et al.*

There are three main points of difference between the two sets of results, the most important being that according to Conversi *et al.* the efficiency reaches a maximum at a pressure of neon of 35 cm Hg, whereas the observations reported here show an increase in efficiency up to the highest pressure used, 67 cm Hg. Conversi *et al.* ascribe the reduction in efficiency at higher pressures (45 cm Hg) to the greater probability of electron recombination and increased absorption of the excitation quanta. It should be pointed out, however, that at the two highest pressures in Conversi's work the layer efficiencies are high—$76 \cdot 5 \pm 2 \cdot 5\%$ and $72 \pm 3\%$— and these points are not inconsistent with a constant efficiency above 35 cm Hg or even one increasing slightly with pressure. That these efficiencies are significantly below the maximum attainable value of 85% is attributable partly to the loss of electrons during the time delay, and partly to accidental coincidences of the two-fold counter telescope. The results of the two experiments may not, therefore, be inconsistent.

The next point of difference is in the shape of the (efficiency, field strength) curves. In the work of Conversi *et al.* the initial rise of the curves occurs at lower field strengths than in our experiments. The characteristics have a definite plateau, contrary to our observations which indicate that the efficiencies in general increase continuously with field strength. This discrepancy may result from the different methods of applying the high voltage pulse in the two cases. Conversi *et al.* applied the pulse across four layers of tubes, and consequently the flashing of one tube in a layer might increase the effective field strength on tubes in the remaining layers and so increase the probability of other tubes flashing. Because of the possibility of this effect, the pulse was applied to single layers only in the present work.

Finally, it was found possible to use commercial neon in the tubes. In a private communication (1956) Conversi stated that to achieve a reasonable efficiency with commercial neon a very short time delay, $\lesssim 0 \cdot 3\,\mu$sec, was necessary. With the higher pressures used in this experiment, the impurities in Ne(C) are not important in reducing the efficiency and much longer time delays can be tolerated.

§ 5. CONCLUSIONS

It has been found possible to produce a flash tube of high efficiency which can be mass-produced cheaply. For example, a tube constructed of soda glass, filled with Ne(C) at about 65 cm Hg pressure, having a residual air pressure of about 10^{-3} mm of Hg, has an internal efficiency of about 86% with $T_D = 3 \cdot 7\,\mu$sec, $T_R = 0 \cdot 5\,\mu$sec, $\tau = 4\,\mu$sec and an applied field strength of about 11 kv cm.$^{-1}$ The dead-time of the tube is about 1·5 sec, and the resolving time may be found from the fact that the time delay for the efficiency to fall to 50% of its value at $T_D = 0$ is about 12 μsec. Both these values can presumably be reduced if necessary by applying a static clearing field of the order of 1 kv cm^{-1}, but the efficiency at the normal working value of T_D would then be somewhat reduced.

ACKNOWLEDGMENTS

The authors wish to thank Professor G. D. Rochester for his interest in the work and for many helpful discussions.

We are greatly indebted to Professor M. Conversi for communicating his results to us before publication and to Professors J. D. Craggs and J. M. Meek and Drs. A. von Engel and W. A. Prowse for useful discussions.

One of us (M.G.) wishes to thank the Department of Scientific and Industrial Research for a maintenance allowance, and another (E.R.), the British Council for a Fellowship held partly at the University of Manchester and partly at the Durham Colleges of the University of Durham.

This Research has been supported in part by the Department of Scientific and Industrial Research.

REFERENCES

Barsanti, G., Conversi, M., Focardi, S., Murtas, G. P., Rubbia, C., and Torelli, G., 1956, *Proceedings of C.E.R.N. Symposium*, **II**, 56.

Biondi, M. A., and Brown, S. C., 1949, *Phys. Rev.*, **76**, 1697.

Conversi, M., Focardi, S., Franzinetti, C., Gozzini, A., and Murtas, P., 1955, *Proceedings of the Pisa Conference on Elementary Particles*, 59 (*Nuovo Cim.* Suppl., in the press).

Conversi, M., and Gozzini, A., 1955, *Nuovo. Cim.*, **2**, 189.

Druyvesteyn, M. J., and Penning, F. M., 1940, *Rev. Mod. Phys.*, **12**, 87.

Eyeions, D. A., Owen, B. G., Price, B. T., and Wilson, J. G., 1955, *Proc. Phys. Soc.* A, **68**, 793.

Healey, R. H., and Reed, J. W., 1941, *The Behaviour of Slow Electrons in Gases* (Sydney: Amalgamated Wireless of Australasia).

Holmes, J. E. R., Owen, B. G., Rodgers, A. L., and Wilson, J. G., 1956, *Report of the Conference on Recent Developments in Cloud Chamber and Associated Techniques*, March 1955. Eds N. Morris and J. Duff (London: University College), p. 177.

Hyams, B. D., Mylroi, M. G., Owen, B. G., and Wilson, J. G., 1950, *Proc. Phys. Soc.* A, **63**, 1053.

Jesse, W. P., and Sadauskis, J., 1952, *Phys. Rev.*, **88**, 417.

Kruithof, A. A., and Penning, F. M., 1937, *Physica*, **4**, 430.

Littleton, J. T., and Morey, G. W., 1933, *The Electrical Properties of Glass* (New York: Wiley).

Pidd, R. W., and Madansky, L., 1949, *Phys. Rev.*, **75**, 1175.

Robinson, E., 1953, *Proc. Phys. Soc.* A, **66**, 73.

Rossi, B., and Staub, H. H., 1949, *Ionisation Chambers and Counters* (New York: McGraw Hill).

RESEARCH NOTES

The Lattice Constants of the Inert Elements

By G. H. CHEESMAN and C. M. SOANE
University of Tasmania, Hobart, Tasmania

MS. received 19th March 1957

Considerable interest attaches to measurements of the lattice constants of solidified inert gases at low temperatures because of the opportunity in this relatively simple case of deducing information on the form of the law of force between molecules. Nevertheless, techniques for the measurements are not easy and the published observations, mostly at isolated temperatures, show inconsistencies among themselves. In table 1 are collected together the results given by Simon and Ruhemann (1932) (SR), and by Natta and Nasini (1930) (NN) for xenon and krypton, and by Keesom and Mooy (1930) (KM) for krypton, and the scatter may be judged.

Table 1. Published Values of Lattice Constants for Xenon and Krypton

Substance	Temp. (°K)	Lattice constant (kx)	Authors
Xenon	88	$6{\cdot}24 \pm 0{\cdot}025$	SR
	100	$6{\cdot}18 \pm 0{\cdot}01$	NN
Krypton	20	5·59	KM
	78	5·78	NN
	82	$5{\cdot}68_4 \pm 0{\cdot}02$	SR
	89	$5{\cdot}69_4 \pm 0{\cdot}01_7$	SR
	92	$5{\cdot}69_7 \pm 0{\cdot}01_4$	SR

During the years 1949 to 1952 the writers developed an apparatus capable of giving lattice constants of considerable precision at temperatures down to about 20°K and set out with this to investigate the elements Xe, Kr, Ar and Ne each over as long a temperature range as possible down from the triple point. Unfortunately they dispersed before the programme could be completed, and as it now seems unlikely that further work will be possible for a considerable time, the results already available are presented in view of current interest in this field. A more detailed account of the apparatus, which has wide applications, will be presented elsewhere.

The fundamental unit of the apparatus consisted of two cylindrical vessels for refrigerant arranged coaxially end to end; the contiguous ends of these vessels were concave and thus enclosed a lens-shaped volume between them, about 7 cm in diameter and 1 cm thick at the axis, diminishing to a narrow slit about 2 mm wide at the circumference. Along the axis of this space was a wire of copper (diameter 0·037 cm, 28 s.w.g.) which supported the specimen. This wire also provided a reference for measurement by superposing the diffraction lines of copper upon the pattern from the specimen, and in addition it formed one element of a thermocouple by which the temperature of the specimen was measured. (An attempt was made to use a platinum supporting wire which would also serve as a resistance thermometer, but its thermal conductance was not large enough to enable good deposits to be obtained on a fine wire.) Since the angle subtended by the circumferential slit at the specimen was only a few degrees, access of heat by radiation from the surroundings was kept very small.

This specimen unit was suspended rigidly in another double-walled vessel which also had a narrow gap in the plane of the specimen chamber, and the space between its walls could be filled with an appropriate refrigerant. This vessel was found very valuable in securing steadiness of temperatures in the innermost vessel, and it was essential when liquid hydrogen was used. Beyond the outer refrigerant vessel was a polished radiation shield, and beyond that a metal ring supporting a strip of film at a radius of 19 cm from the specimen. The Bradley–Jay arrangement was used, and a series of notches in the edge of the film ring provided fiducial marks on the film as a check on any movement which might occur as a result of temperature changes. The whole assembly was placed in a vacuum-tight container fitted with an aluminium window through which was introduced a fine pencil of x-rays (unfiltered copper radiation); a second porthole carried a retractable glass nozzle through which the gas under examination could be blown on to the central wire to form the solid specimen. Exposures of 8 to 10 hours, using a commercial CA-6 type of tube, gave satisfactory patterns.

Commercial gases, stated to be of 99–100% purity (B.O.C.) were used without further treatment. The principal impurity in the Xe was stated to be Kr, and preferential condensation should reduce the proportion of Kr in the solid. The Kr was stated to contain Xe and the proportion might be increased in the solid. No lines of Xe were found on the Kr photographs, but solid solutions with an increased lattice constant might possibly be present.

The photographs were measured in the conventional manner and the diffraction angles calculated from the linear measurements by conversion factors interpolated graphically from the copper lines on the same film, thereby eliminating eccentricity errors. Between 12 and 18 inert-element lines were measured on each film, and the calibration was based on a similar number of copper lines in the same angular range. After indexing the xenon or krypton lines individual lattice constants were calculated and plotted against Nelson and Riley's functions

$$\tfrac{1}{2}\left\{\frac{\cos^2\theta}{\theta}+\frac{\cos^2\theta}{\sin\theta}\right\}.$$

The 'best' straight line through these points was drawn and extrapolated to $\pi/2$ in order to obtain the final lattice constant. In table 2 are shown the results from ten photographs; the limits shown in each are the extreme deviations of single points on the plot, and we consider the limits given for the means are a conservative estimate of the overall precision. The temperatures are reliable to a few tenths of a degree.

Table 2. Lattice Constants of Xenon and Krypton, referred to Cu = 6·2150 kx at 20°c

Substance	Temp. (°K)	Lattice constant (*k*x)	Mean (*k*x)
Xenon	58	6·186 ± 0·002	
		6·185 ± 0·002	
		6·185 ± 0·004	6·185 ± 0·002
	77·4	6·208 ± 0·002	
		6·205 ± 0·002	
		6·203 ± 0·002	
		6·208 ± 0·002	6·206 ± 0·003
	88·9	6·216 ± 0·004	
Krypton	58	5·710 ± 0·002	
		5·711 ± 0·003	5·710 ± 0·003
	68	5·734 ± 0·01	

These measurements were made at the University of Reading and the authors are grateful to Professor E. A. Guggenheim for facilities for building the camera. One of us (C.M.S.) is also grateful to the Department of Scientific and Industrial Research for a grant.

REFERENCES

KEESOM, W. H., and MOOY, H. H., 1930, *Nature, Lond.*, **125**, 889.
NATTA, G., and NASINI, A. G., 1930, *Nature, Lond.*, **125**, 459, 889.
SIMON, F. E., and RUHEMANN, M., 1932, *Z. Phys. Chem.* B, **15**, 389.

Dielectric Constant and Conductivity of Guanidine Aluminium Sulphate during Switching

BY M. PRUTTON
Birkbeck College Computational Laboratory, University of London

MS. received 1st *April* 1957

IN the course of work on the domain structure of guanidine aluminium sulphate hexahydrate (GASH), some experiments were performed to study the variation of the dielectric constant and the equivalent parallel resistance as the crystals were being switched at low frequencies.

Crystals of GASH grown by evaporation and in a rotary crystallizer were cleaved into small plate-like samples about 0·02 cm thick, and gold electrodes of area about 0·5 cm^2 were evaporated on to them. The impedance of these crystals was measured by using a bridge technique at frequencies between 50 and 200 kc/s at the time they were being switched by a 50 c/s sine wave (Drougard, Funk and Young 1954). The measuring voltage amplitude was kept very much less than the switching voltage amplitude so that the switching mechanism of the crystal was not affected. The frequency limits were set by practical difficulties in the apparatus and not by the material. The impedance was measured at the peak value of the switching current density i_e, and so the dielectric constant ϵ_s and the equivalent high frequency conductivity σ_s were found as functions of the measuring frequency f and of i_e.

As in the case of barium titanate there is a marked increase in the loss component of high frequency current flowing through the crystal while switching occurs. The equivalent conductivity was found to be given by $\sigma_s = ai_e/E_c^2$ where E_c is the coercive field strength and a the activation field which arises in the expression

$$\frac{dP}{dt} = F(P)\exp\left\{-\frac{a}{E(t)}\right\}$$

found to be true in previous pulse experiments carried out on GASH.

Drougard, Funk and Young (1954) assume that the coercivity of barium titanate is a constant during switching but this is not the case for GASH as its coercive field strength was found to vary rapidly with the peak amplitude of the switching wave; also the coercivity was found to be a function of frequency (figure 1).

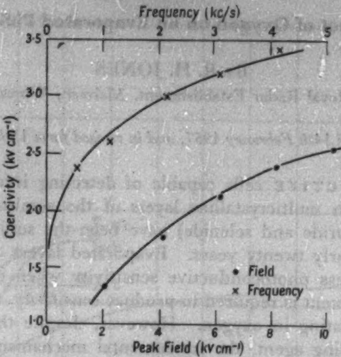


Figure 1. **Coercivity versus peak switching field (50 c/s) and switching frequency.**

Values of a found by plotting σ_s against i_e/E_c^2 were about 4000 v cm^{-1} which figure agrees reasonably well with the values of a found from the pulsing experiments. The variation of σ_s and ϵ_s with i_e and f is shown in figures 2 and 3.

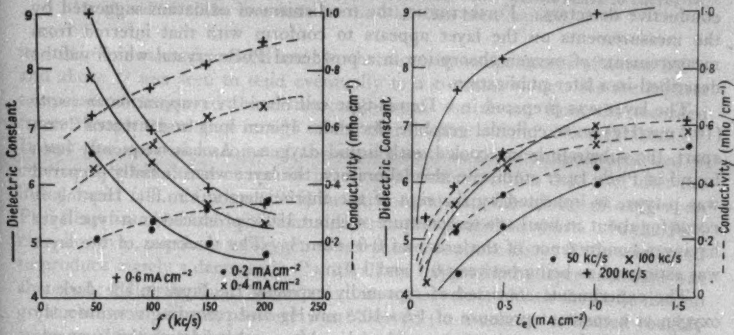


Figure 2. Dielectric constant and conductivity versus measuring frequency.

Figure 3. Dielectric constant and conductivity versus switching current density.

It is surprising that the dielectric constant increases on switching: since the current flow due to the domain formation reaches a maximum after the switching voltage reaches the coercive value (i.e. the switching current lags the switching voltage) we would expect to see an extra inductive component impedance on applying the switching wave and not a capacitative component. Furthermore, the increase of ϵ_s for barium titanate found by Drougard, Funk and Young was proportional to i_e but this is not the case for GASH. The results do fit a relaxation type of spectrum for a dielectric constant of 5 measured in the absence of a switching voltage and a relaxation time of about 40 microseconds. The behaviour of ϵ_s remains essentially unexplained.

REFERENCES

DROUGARD, M., FUNK, H., and YOUNG, D., 1954, *J. Appl. Phys.*, **25**, 1166.

The Effect of Oxygen on an Evaporated PbSe Layer

By R. H. JONES
Royal Radar Establishment, Malvern, Worcs.

MS. received 14th February 1957, and in revised form 15th April 1957

PHOTOCONDUCTIVE cells capable of detecting infra-red radiation and employing thin multicrystalline layers of the semiconducting lead salts (sulphide, telluride and selenide) have been the subject of research and development for nearly twenty years. Evaporated layers of all three salts do not in general possess photoconductive sensitivity when initially prepared *in vacuo*; further treatment is required to produce sensitivity, the most effective so far known being heating in oxygen. However, despite the continued use of oxygen as a sensitizing agent, the fundamental mechanism of the sensitizing process is not yet fully understood. The present investigation has been undertaken with a view to obtaining an understanding of this mechanism in PbSe layers. Preliminary results, which are now reported, were obtained in one layer only, but it is believed that the qualitative behaviour observed is typical of that obtained with a large number of the layers processed in this laboratory as photoconductive detectors. Furthermore, the mechanism of oxidation suggested by the measurements on the layer appears to conform with that inferred from measurements of oxygen absorption in a powdered PbSe crystal which will be described in a later publication.

The layer was prepared in a Dewar-type cell blank by evaporation *in vacuo* (10^{-6} mm Hg) on to colloidal graphite electrodes 16 mm long and situated 2 mm apart, the substrate being cooled with liquid oxygen. As has frequently been found in PbSe layer studies in this laboratory, the layer when initially prepared was p-type, as indicated by the sign of the thermoelectric e.m.f. Heating *in vacuo* for about an hour at a temperature of about 350°C produced an n-type layer having a conductance of the order of 0·01 ohm^{-1}. The thickness of the layer was estimated as being between 0·1 and 1·0 μ.

The experiments consisted of repeatedly exposing the layer in the dark to oxygen at a constant pressure of $1{\cdot}6 \times 10^{-2}$ mm Hg and recording continuously the variation in layer conductance produced. After each oxidation the layer was heated *in vacuo* at about 350°C for an hour, this process being found to restore the layer to approximately the same initial conductance each time ($0{\cdot}005 \pm 30\%$ ohm^{-1}). Merely evacuating the cell at room temperature did not appear to alter the layer conductance even after exposure to oxygen at the same low temperature.

Figure 1 shows the variation with time of layer conductance C obtained from a typical set of successive oxidations representing a total of twenty-five carried out at fixed temperatures T in the range from room temperature to about 340°C. For values of T from room temperature to nearly the highest employed, exposure to oxygen produced a sharp fall in C followed by a much less rapid rise. In all cases this process transferred the layer from n-type to p-type, as determined by the sign of the thermoelectric e.m.f. obtained at room temperature. It appeared that a similar behaviour was obtained at 0°C also, though in this case the rate of change of C was slow and the variation in C was not studied beyond the minimum conductance stage. With rise of temperature the time required for C to reach

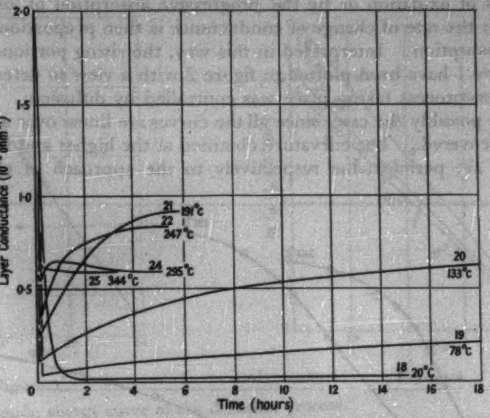


Figure 1. Change in layer conductance at various temperatures produced by exposure to $1{\cdot}6\times10^{-2}$ mm pressure of oxygen. The time scale shown refers to curves 18, 19, 20 and 21. The time scale for the others has been doubled.

its minimum value decreased and, furthermore, at temperatures of about 200°c and above *C* was seen to tend eventually to a constant value. At the highest temperatures employed the distinct minimum was absent from the recorded traces though the change in sign of the thermoelectric e.m.f. was still obtained. It was also noticed in some of the experiments at these temperatures that *C* continued to fall slowly or remain practically constant, rather than rise, after the initial rapid fall. Minden (1955) in experiments on the effects of oxygen on PbS films has observed behaviour similar to that described above. In the PbS case, however, exposure to oxygen at temperatures above about 200°c was found to produce merely a decrease in *C* until all the oxygen in the layer enclosure was absorbed; following this, *C* returned to its initial value. With the PbSe film no great change in oxygen pressure was observed (in a volume estimated roughly as being 200 cm³), and there was no indication of *C* reverting eventually to its original value. These experiments, however, were carried out independently of those of Minden, and this possibility was not specifically examined.

If, as seems likely, the PbSe layer material at the beginning of each oxidation possessed an excess of lead, being n-type, the initial decrease in layer conductance could be attributed to the filling up of selenium vacancies in the crystallite lattices, the minimum conductance being attained when an approximately balanced impurity state was reached. Further absorption of oxygen in producing positive holes would cause the conductance to increase. In these experiments the initial fall in *C* occurred, in general, too rapidly to enable the variation in *C* at this stage to be followed accurately. The subsequent increase, however, taking place less rapidly, was recorded clearly at all but the highest temperatures employed. Now such results, expressed as the variation of conductance with time, can be used to investigate the rate of uptake of oxygen, provided that the mobility of the current carriers is not altered greatly by changes in the fixed temperatures of oxidation or by the progressive absorption of oxygen at each temperature; the rate of change of conductance is then proportional to the rate of oxygen absorption. Interpreted in this way, the rising portions of curves 18 to 22 of figure 1 have been plotted in figure 2 with a view to determining if the gas absorption process taking place was controlled by diffusion. This appears, in fact, to be possibly the case, since all the curves are linear over a part of or the whole range covered. The curvature obtained at the higher and lower values of conductance are perhaps due respectively to the approach of chemical equilibrium in the layer, and mixed conduction in the layer. Considering the rates of reaction corresponding to the linear portions of the curves, an estimate has been made of the activation energy of the diffusion process in the temperature range 20 to 247°C, making use of the equation for bulk diffusion (Anderson, Roberts and Harper 1955):

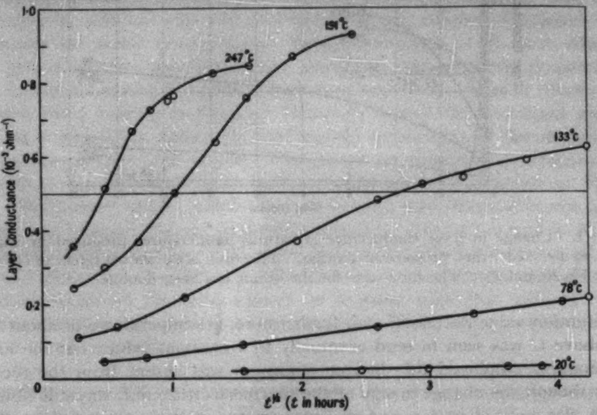


Figure 2. Variation of layer conductance with time after attainment of minimum conductance state.

$$C = kD^{1/2}t^{1/2}$$

where C is the quantity of gas taken up by diffusion in time t, and D is the diffusion coefficient at temperature T, the temperature dependence of D being

$$D = D_0 \exp(-E_D/\boldsymbol{R}T),$$

where $\boldsymbol{R}$ is the gas constant (1·98 cal deg^{-1} mol^{-1}). The plot obtained is shown in figure 3 which gives for E_D the value 14 cal deg^{-1} mol^{-1}.

This value of activation energy, it will be realised, on the basis outlined above, applies only to the process taking place after the filling up of selenium vacancies. The nature of the initial process is not revealed by the layer measurements, but will be discussed in the later publication on oxygen absorption in powdered PbSe.

The condition for minimum conductivity in a mixed semiconductor as described above is also that for intrinsic conductivity provided that the mobilities

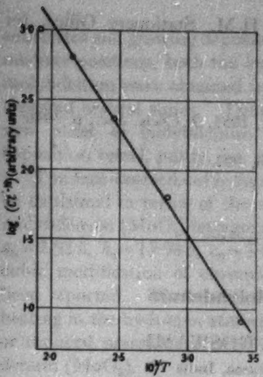


Figure 3.

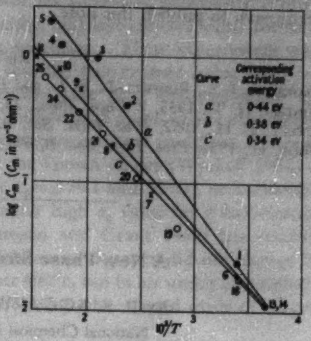


Figure 4.

Figure 3. Activation energy curve of the diffusion process.

Figure 4. Exposure of layer to $1{\cdot}6\times10^{-3}$ mm pressure of oxygen at various temperatures T°K. The variation with T of the minimum value C_m of C attained in three series of experiments. (The number against each plotted point refers to the particular experiment in which the result was obtained.)

of holes and electrons are nearly equal †. It follows, therefore, that subject to this limitation, the values C_m of minimum conductance obtained in these measurements can be used to obtain a value for the activation energy of the conduction process by plotting $\log C_m/f(T)$ against $1/T$, as described by Wood (1956) and previous workers. The function $f(T)$ depends on the temperature variation of carrier mobility. This variation is not known accurately for layers, though measurements (Putley, private communication) indicate that this is slower than that given by the $T^{-5/2}$ law found for single crystal PbSe (Putley 1952). Assuming, therefore, that the effect of $f(T)$ is small in comparison with the exponential term in T in the equation for the conductivity of the semiconducting film, the results of three successive series of experiments have been plotted in the form ($\log C_m$, T^{-1}) as shown in figure 4, yielding the respective values 0·44, 0·38 and 0·34 ev for the activation energy. These are comparable with the mean value of about 0·37ev for the energy gap in PbSe obtained by Putley (private communication) from Hall and conductivity data on single crystals in the same temperature range.

The suggestion from this observation is that interaction of oxygen with the layer did not produce a marked change in the band structure of the layer material, but instead provided acceptor centres for electrons. The reason for the progressive decrease in the slopes of the lines in figure 4 is not yet clear. This change of slope may be taken as indicating that the prolonged series of oxygen admissions and withdrawals had some permanent effect on the layer material band structure, possibly by the gradual formation of a new stable compound, though, on the other hand, one cannot exclude the possibility that the change was produced at least to some extent by changes in the physical structure of the layer.

† Recent measurements carried out by the author on PbSe single crystals indicate that in the temperature range from room temperature to 300°c, the mobility ratio is between about 1 and 2.

Acknowledgment is made to the Controller, H.M. Stationery Office, for permission to publish this note.

References

Anderson, J. S., Roberts, L. E. J., and Harper, E. A., 1955, *J. Chem. Soc.*, p. 3946.
Minden, H. T., 1955, *J. Chem. Phys.*, **23**, 1948.
Putley, E. H., 1952, *Proc. Phys. Soc.* B, **65**, 388, 736, 992.
Wood, C., 1956, *Proc. Phys. Soc.* B, **69**, 613.

A New Phase Structure of Molybdenum

By P. S. AGGARWAL and A. GOSWAMI
National Chemical Laboratory, Poona, India

Communicated by A. B. Biswas ; *MS. received 23rd October* 1956, *and in revised form 25th April* 1957

Polymorphism is exhibited by many metals. Examples are the α–γ modifications of iron (Taylor 1952), face-centred cubic and hexagonal close-packed forms of cobalt, nickel, silver and copper (Taylor 1952, Finch, Wilman and Yang 1947, Layton 1952) and the body-centred cubic, hexagonal close-packed and face-centred cubic structures of chromium (Taylor 1952, Abrahamson and Grant 1956). The body-centred cubic form of nickel (Finch, Sinha and Goswami 1955), on annealing at about 400°c, reverted to the normal face-centred cubic structure.

Molybdenum has a body-centred cubic structure and no face-centred cubic modification has as yet been reported. We have, however, observed that molybdenum deposited *in vacuo* often develops a face-centred cubic structure.

A pure molydbenum wire (supplied by Messrs Tungsten Manufacturing Co., London) was used as a filament for our experimental purpose. It was initially flashed for several seconds to remove the surface impurities. The deposition was carried out for about 10 minutes to half an hour or more *in vacuo* ($\simeq 10^{-4}$ to 10^{-5} mm Hg) on to both rocksalt and glass substrates situated about 2 cm away from the filament.

The electron diffraction pattern (figure 1, Plate) shows that the deposits formed on rocksalt in about 10 minutes were in two-degree {100} and {111} orientations. The pattern has spots or rings, the *hkl* values of which were all odd or all even. In conjunction with the relative ring intensities, this points to the structure being face-centred cubic. Even when the deposits were so thick as to be opaque (several thousand ångströms thick or more) due to a longer time of deposition, this structure was retained and the crystals were one-degree {100} (figure 2, Plate) or {111} orientated. Similar orientations were also observed from thick deposits formed on glass substrates. The lattice parameter a_0 (referred to $d_{11\bar{2}0}$ of graphite $= 1{\cdot}230$ Å) was found to be $4{\cdot}16$ Å. If, on the other hand, the substrates were close to the filament during deposition, molybdenum crystals of the body-centred cubic structure with normal or tilted {111} orientation were formed, no doubt, as a result of the heating of the substrates and growing deposits. If, however, the substrates were placed at intermediate positions, both the body-centred cubic and face-centred cubic forms of molybdenum were obtained from the same specimen. These experiments were repeated several times. Thus it appears that the face-centred cubic form is a new phase of molybdenum. Similar results were also obtained when the deposition vessel, pump, etc. were initially flushed with argon before evacuating.

The face-centred cubic form with a high lattice constant could not, however, be explained in terms of the structures of the known oxides (Wyckoff 1951) of molybdenum (MoO_2, tetragonal $a_0 = 4{\cdot}44$ Å, $c_0 = 2{\cdot}79$ Å and MoO_3, orthorhombic $a_0 = 3{\cdot}92$ Å, $b_0 = 13{\cdot}94$ Å, $c_0 = 3{\cdot}66$ Å). A similar high a_0 value for a face-centred cubic modification of chromium (Abrahamson and Grant 1956) has recently been reported. The structure of the molybdenum deposits did not change by heating in the hydrogen atmosphere at about 400°c, but in air under atmospheric or reduced pressures these films were converted to the normal oxide of molybdenum (MoO_3).

The mechanism of this phase formation is not clearly understood. It is generally known that both in martensitic and non-martensitic types of transformations, the orientation relationships are similar for similar structures (see Bowles and Barrett 1952). In the α–γ transformations of iron, different mechanism (Kurdjumow and Sachs 1930, Sachs 1932, Nishiyama 1934–5) has been proposed. Agarwala and Wilman (1953, 1954) have recently suggested an alternative way for the phase change which occurs on the $\{110\}/\{100\}$ planes in such a way that $\langle 111 \rangle$ axis of the former is changed into $\langle 110 \rangle$ of the latter and vice versa. Whatever may be the mechanism, in an ideal case, the face-centred cubic phase should have $a_0 = \sqrt{(3/2)}\,(= 1{\cdot}225)$ that of the corresponding body-centred cubic value. The lattice parameters of a few polymorphic metals are set out in the table. In those cases where the face-centred cubic parameters are not known from direct experiment they have been calculated from those of the known hexagonal close-packed values, the relation being $a_{0(\mathrm{fcc})} = a_{0(\mathrm{hcp})}\sqrt{2}$. It will be seen that in all cases, the ratio somewhat exceeds the ideal value particularly more so in the case of chromium and molybdenum. Such an expansion of lattices is not unlikely during the readjustment of atoms in the thermal transformation process. A detailed account will be published later on.

Polymorphic Structures of Metals

Metal	a_0 (fcc)	a_0' (bcc) ‡	a_0 (hcp)	c_0 (hcp)	c_0/a_0	Observed a_0/a_0' (calc. 1·225)
Ca	5·560	4·477	3·94	6·46	1·639	1·24
Fe	3·564	2·861	—	—	—	1·25
Tl	†4·878	3·874	3·450	5·514	1·598	1·26
Zr	†4·558	3·61	3·223	5·123	1·590	1·26
Ni	3·517	2·78	2·49	4·08	1·64	1·27
Cr	†3·841	2·88	2·717	4·418	1·626	1·33
	3·68	2·88	—	—	—	1·28
	3·98	2·88	—	—	—	1·38
Mo	4·16	3·141	—	—	—	1·32

† Calculated from $a_{0(\mathrm{hcp})}$ (hypothetical)
‡ Smith, Carlson and West 1956.

Acknowledgment

Our thanks are due to Professor G. I. Finch and Dr. A. B. Biswas for their keen interest. We are also grateful to Dr. M. K. Gharpurey for the help during some of our experiments.

References

Abrahamson, E. P., and Grant, N. J., 1956, *Trans. Amer. Inst. Min.* (*Metall.*) *Engrs*, **8**, 975.

Agarwala, R. P., and Wilman, H., 1953, *Proc. Phys. Soc.* B, **64**, 717 ; 1954, *Proc. Roy. Soc.* A, **223**, 167.

Bowles, J. S., and Barrett, C. S., 1952, *Progress in Metal Physics* (ed. B. Chalmers), Chap. I, Vol 3 (London : Pergamon Press).

Finch, G. I., Sinha, K. P., and Goswami, A., 1955, *J. Appl. Phys.*, **26**, 250.

Finch, G. I., Wilman, H., and Yang, L., 1947, *Disc. Faraday Soc.*, **1**, 144.

Kurdjumow, G., and Sachs, G., 1930, *Z. Phys.*, **64**, 325.

Layton, D. N., 1952, *J. Electrodep. Tech. Soc.*, **28**, Adv. copy No. 9, 6.

Nishiyama, Z., 1934–5, *Sci. Rep. Tohoku Univ.*, **23**, 638.

Sachs, G., 1932, *Z. Metallkde*, **24**, 241.

Smith, J. F., Carlson, O. N., and West, R. W., 1956, *J. Electrochem Soc.*, **103**, 409.

Taylor, A., 1952, *An Introduction to X-Ray Metallography* (London : Chapman and Hall), pp. 118, 382.

Wyckoff, R. W. G., 1951, *Crystal Structures* (New York : Interscience Publishers), Vol. I.

LETTERS TO THE EDITOR

Comments on the Manuel Model for MnBi

Recently Manuel (1956) has proposed a model for the interpretation of the magnetic properties of MnBi and MnAs. He suggests that below the Curie temperature T_C these compounds have a three-dimensional ferromagnetic lattice, while above T_C two-dimensional ferromagnetic planes occur. These ferromagnetic planes are perpendicular to the c axis and neighbouring planes are oriented at random. Since the ferromagnetic sheets are only capable of having their moments directed in the plane perpendicular to the c axis, the Curie constant can easily be shown to be equal to $C = N\mu^2/2k$, where N is the number of ferromagnetic units per mol of Mn atoms, μ is the moment of a ferromagnetic unit. If n is the number of Mn atoms in a ferromagnetic unit, there are $N = A/n$ (A = Avogadro's number) units in a mol, each having a moment $\mu^2 = \mu_B{}^2 g^2 nJ\,(nJ+1) \simeq (\mu_B gnJ)^2$. Therefore $C = Ag^2nJ^2\mu_B{}^2/2k$. Assuming each Mn^{3+} contributes $4\mu_B$ (as in MnBi) we find $C = 2{\cdot}98n$. Thus one can see that even ferromagnetic sheets containing as few as 10 atoms would give $C = 29{\cdot}8$ which is more than a factor of 10 larger than the experimental value of 2·3 found for MnBi (Heikes 1955), and 2·4 found for MnAs (Guillaud 1951).

In addition, a recent neutron diffraction study (Roberts 1956) has shown that a disordered arrangement of Mn atoms on regular interstitial lattice sites most closely fits the observed data above the Curie temperature. The disordered model qualitatively explains several of the unusual properties of MnBi. The movement of about 10% of Mn atoms to large interstitial sites would explain the 3% contraction of the c_0 axis observed upon loss of magnetization. Also, the observed hysteresis can be justified by assuming a distortion energy to be associated with the movement of an Mn atom to an interstitial site.

It thus appears that the predictions of Manuel's model are inconsistent with the observed paramagnetic susceptibility data for MnBi and MnAs. In addition neutron diffraction data for MnBi do not bear out his model.

Westinghouse Research Laboratories,
Pittsburgh, Pa.
R. R. Heikes.
General Electric Research Laboratories,
Schenectady, New York.
B. W. Roberts.

18th February 1957, in revised form 24th April 1957.

Guillaud, C., 1951, *Grenoble Conference Papers*, p. 75.
Heikes, R. R., 1955, *Phys. Rev.*, **99**, 446.
Manuel, A. J., 1956, *Proc. Phys. Soc.* B, **69**, 691.
Roberts, B. W., 1956, *Phys. Rev.*, **104**, 607.

REVIEWS OF BOOKS

Proceedings of the Second International Congress of the International Union for the Philosophy of Science: Zurich, 1954. (Neuchâtel/Suisse: Editions du Griffon; Paris: Dunod, 1955). 5 vols. (pp. 159, 152, 170, 136, 107). 3,000 *fr*.

Five volumes have been published of the proceedings of this Congress which was convened by Swiss philosophers of science. These volumes are devoted to (1) *Plenary Sessions*, (2) *Physics and Mathematics*, (3) *Theory of Knowledge and Linguistics*, (4) *Philosophy and Science*, *History of Philosophy*, (5) *Sociology and Psychology*. The printing was done in Switzerland and is very neat and clear.

Unfortunately, the papers themselves form a very curious hotch-potch, reflecting the equally curious way in which this Congress was organized. Some are of considerable interest, but many are of little value and others quite out of place. There seems, too, to be very little sense of arrangement. For example, Adolf Grünbaum's mathematical paper on the clock paradox in the special theory of relativity appears in the third volume under Theory of Knowledge, whereas the abstract of Pierre Vernotte's paper *Les principes de la méthode scientifique* is included in the second volume under Physics.

Among the comparatively few papers which I can strongly recommend to readers of these *Proceedings* are the review of *The Philosophy of Science of Logical Empiricism* by Herbert Feigl and Piaget's paper on *Les lignes générales de l'epistemologie génétique*, both in the first volume, the paper by E. H. Hutten *On the Pauli Principle* in the second volume, and the paper on the clock paradox, already referred to, in the third volume.

Feigl's paper is the most valuable. It is a lucid discussion of some of the more important recent changes and developments in the outlook of Logical Empiricism. Feigl distinguishes three main trends. The first, exemplified in the work of Philipp Frank, combines informal logical analyses of the sciences with awareness of the psychological and cultural factors which lead to the selection of problems and may be regarded as the sequel to the work of Ernst Mach. Second is the development of analytic philosophy by G. E. Moore and Wittgenstein (to the reviewer this development seems to have very little to do with science and justifies Gustav Bergmann's appelation of 'futilitarian'). The third method is exemplified in the work of Carnap, Reichenbach, Woodger, Braithwaite, Tarski and others. Feigl argues incisively against Quine's contention that there is no sharp distinction between analytic and synthetic statements. Although the reviewer tends to sympathize with Quine in principle, he found Feigl's discussion clear and cogent, e.g. " The applicability of a certain formal structure such as transitivity to a set of empirically given facts is clearly a matter for synthetic statements. And although it might be said that Newtonian time is analytically transitive, or that Euclidean triangles are analytically equiangular, it cannot be claimed that the transitivity of experienced (or physically measured) time is a matter of analytic truth. Nor can it be said that it is logically true that a physical triangle *empirically* determined as equilateral must therefore also be equiangular." Furthermore, his discussion of Duhem's view of the impossibility of crucial experiments in science contains much sound sense, and his serious doubts concerning the role of alternative multi-valued logics as a solvent of scientific problems seem to be well-founded. All in all, a very stimulating paper.

G. J. WHITROW.

Acht- und Neunstellige Tabellen zu den Elliptischen Funktionen, by M. SCHULER and H. GEBELEIN. Pp. xxiv + 296 (Berlin: Springer, 1955). DM58.

This volume represents another attempt to tabulate the elliptic functions in a way which will enable them to be used more easily in numerical calculations. Theory introduces them, in the most elementary way, as generalizations of the circular functions, sn u being the inverse of the familiar elliptic integral. In any case, the functions sn u, cn u, dn u, are inconvenient to tabulate. They are functions of two variables, and have awkwardly placed infinities.

In the present volume, the tables of most direct use to the computer give the logarithms of sn $u/\sin x$, cn $u/\cos x$ and of dn u as functions of z. There are also tables from which the elliptic integrals K and E may be obtained directly. Here $x=\pi u/2K$ and $z=\cos 2x$. The modulus is taken as Jacobi's q, and the relation between this and Legendre's θ is also tabulated.

The selection of this form in which to tabulate the functions appears likely to be a useful one, but only longer experience can really show whether this is so.

The book is well printed and bound, and the explanation of the tables is given in English as well as German.

J. H. AWBERY.

Fünfstellige Tabellen zu den Elliptischen Funktionen, by M. SCHULER and H. GEBELEIN. Pp. x + 114 (Berlin: Springer, 1955). DM 29·60.

The main tables are those of $\log \operatorname{sn} u/\sin x$, $\log \operatorname{cn} u/\cos x$, and $\log \operatorname{dn} u$, with the modulus taken as Jacobi's q and not Legendre's $k=\sin\theta$, but a table of the relation between q and θ is given. The arguments u and x are related by $u=2kx/\pi$ where, of course, K is the complete integral of the first kind for the modulus q, which again is tabulated separately.

The quantities mentioned in the first line are actually given as functions of $z=\cos 2x$ and q, so that quite a lot of auxiliary calculation is necessary before the value of sn (u, k) can be obtained from the tables. However, all aid needed for interpolation is supplied, and only experience can show whether these tables will rob computation with elliptic functions of the tedium hitherto associated with it.

J. H. AWBERY.

Thermal Power from Nuclear Reactors, by A. S. THOMPSON and O. E. RODGERS. Pp. xiii + 229. (New York: John Wiley; London: Chapman and Hall, 1956.) 58*s*.

This book, which is based partly upon a series of reactor engineering lectures given by A. Stanley Thompson at the Oak Ridge National Laboratory, is intended primarily for mechanical engineers. The first half of the book reviews the relevant nuclear reactor theory. An introductory chapter describes briefly how a reactor works, and surveys the main reactor types. A concise account is then given of the nuclear principles involved, followed by consideration of the necessary criticality conditions. The reactor kinetic behaviour is discussed at some length, since the authors wisely emphasize the important role of reactivity coefficients in power reactors. The chapter on reactor shielding, however, appears somewhat cursory for a book intended mainly for engineers. In the second half of the book, consideration is given to aspects of reactor technology of more immediate engineering importance. A short but useful chapter on reactor materials is followed by sections discussing thermal stress, power extraction from reactors, and thermal power cycles. Throughout the book, problems are considered by means of such engineering techniques as numerical and dimensional analysis.

The authors have aimed at setting out the text in a manner leading as directly as possible to the problem of using nuclear reactors for the generation of mechanical or electrical power. Perhaps for this reason, they appear to have produced a work which is more in the nature of an introduction to the subject than a self-contained textbook of nuclear engineering. Although the book cannot be unreservedly recommended, therefore, nuclear engineers should find in it much that is good, whilst the first half of the book provides a suitable introduction for any serious student entering the field of reactor technology.

H. ROSE.

Abacs or Nomograms, by A. GIET. Pp. ix + 225. Translated and revised by J. W. Head and H. D. Phippen. (London: Iliffe; New York: Philosophical Library.) 35*s*.

Even in the age of electronic computers, the nomogram will still be welcomed by physical scientists and engineers to help with routine approximate calculations. This book can be recommended for its lucid exposition of the methods by which the common types of nomogram are constructed. Both author and translators are to be congratulated on the way in which the topic is presented. As is necessary, the diagrams are excellent.

On a first reading it seems free from error and lacking in ambiguities. The many examples are carefully laid out in stepwise progression. There is no index, but the table of contents is very comprehensive. There is also an appendix which summarizes the types of abac appropriate to various forms of relationship between the variables to be plotted.

It is very doubtful whether most people would be prepared to pay much of a price to possess a personal copy, since the construction of nomograms is essentially an occasional exercise for which most of us would rely on library books for guidance. The publishers seem to have us weighed up, and have fixed a sufficiently high price to break even. The book should appear in most libraries and, once it is known, I have little doubt it will become the most popular of the reference books on this topic at present available.

M. J. MORONEY.

Electricity, Magnetism and Atomic Physics, Vol. I. *Electricity and Magnetism*, by J. H. FEWKES, and J. YARWOOD. Pp. vii + 768. (London: University Tutorial Press, 1956.) 36*s*.

Many textbooks have been written on the subject of Electricity and Magnetism and, on the appearance of a new one, the reader looks for some special characteristic which makes it worth while. It may be that the book is of the nature of an old story well told, with a fresh lucidity and a new appreciation of its subject, or it may be that a familiar structure has been retained with the addition of modern furnishings to keep the subject alive and abreast of the times. Some works give the impression that the author has learned a little of the subject in order to write a book about it, while others show clearly that the author, having learned much of his subject and having thought much about it, is urged to write. These authors are experienced teachers who have learned how to tell their story lucidly and how to keep their pupils and readers interested by drawing upon modern applications of the subject. They have realized that there is something new to be done in making room in a modern textbook for comparatively new works, but they still emphasize the importance of the fundamental principles of the subject.

The first part of this volume, about three quarters of the whole, is concerned with such familiar parts of the subject as measuring instruments, bridges, alternating current, electrical images and finally Maxwell's equations. These are treated adequately and in an interesting way.

A special feature of this part of the book is the chapter on units, where m.k.s. units receive a longer explanation than is usual in a work of this kind. This is a welcome addition, but the authors have, no doubt rightly, made use, in general, of the absolute system of units.

The last part, after a chapter on gas discharge phenomena and the emission of electrons from solids, is devoted to subjects which are only beginning to be introduced into textbooks. It includes chapters on electronic tubes and the principles of radio communication including an account of ultra-high-frequency oscillators and of properties of the ionosphere. The last chapter contains a brief account of vector analysis with applications of the notation to a number of electrical problems.

The book is written to meet the needs of readers who wish to acquire the extent and standard of knowledge described in university syllabuses for General Degrees and for the Part I Honours course. It goes considerably beyond what is required for the former and to some extent beyond what is necessary for the latter: Part II candidates will find some of the more advanced and more recent work very helpful to them.

Many examples are given at the ends of chapters and some are worked out as illustrations. These serve to test and guide the reader, especially the undergraduate who has to keep examinations in mind.

This is a book to be recommended. It is written by authors who have knowledge and understanding of the needs of physics students and sympathy for their difficulties.

H. T. FLINT.

An Introduction to Crystallography, by F. C. PHILLIPS. Pp. ix+324. 2nd Edn. (London: Longmans, Green, 1956.) 30*s*.

This is the second edition of a book that has proved to be an invaluable aid to student and teacher alike. It first appeared ten years ago and this is ample testing time for any book that claims to be an introduction to a subject that is basic to the understanding of mineralogy and petrology, to x-ray crystallography, pure and applied, and to the wider but increasingly popular field that now has the diffuse title of solid state physics. There is little doubt that the book will enjoy continued popularity, for one sees in it the cumulative experience and the hallmark of sound teaching methods.

The book is almost a reprint of that which first appeared in 1946 and it is only the addition of a chapter dealing with the elementary principles of x-ray diffraction that makes it a new edition. The first part of the book deals with the external symmetry of crystals and crystal morphology in relation to the 32 crystal classes. On Part 1 alone, the student is well served and it can be used as a convenient reference book.

Part 2 maintains the same high standard and provides an excellent treatment of space lattices and the theory of space groups, using the space group notation now accepted and set out in the *New International Tables for Crystallography*. This, one might say, is the connecting link to the further study of X-ray crystallography and the inclusion of the new chapter on X-ray diffraction by crystals bridges this gap. Here again, although the approach is elementary, one is conscious of the touches of a master in the art of teaching a difficult subject without bogging the reader down with detail; and yet sufficient stimulus is provided to encourage the beginner to further reading. Here, however, we find that there is only one book recommended for further study and that deals only with the interpretation of X-ray photographs. It is a pity that the author has not, for the purpose of continuity, suggested at least one of the good general texts on X-ray crystal analysis which are now available. This is a minor blemish that might well be corrected in the next edition of the book; nevertheless, it can be unreservedly recommended to a beginner in this field. C. H. CARLISLE.

Automatic Digital Calculators, by A. D. BOOTH and K. H. V. BOOTH. Pp. ix + 261. (London: Buttersworths Scientific Publications, 1956). 32*s*.

This book is intended to serve as a guide to the theory of automatic digital computers, and covers both the design of the machines and the methods of application.

The book covers a very wide field and the stress, in the design section of the book, tends to be directed more to a description of the computing units in the form of block diagrams as opposed to a description of the actual circuits themselves.

The last five chapters deal with machines from the users' point of view, and hence will be of most interest to the general reader. Here the authors start by discussing the basic operations necessary for a complete instruction code for a machine, particular reference being made to the authors' own machine, the APE(X)C at Birkbeck College. The next three chapters describe the technique of programming and coding, i.e. the breaking down of the calculation into convenient parts and the translation of these parts into the detailed arithmetical operations which the machine performs. Some numerical methods that might be employed in this connection, for example, interpolation and Chebyshev approximation, are described.

The last chapter describes some applications of computers. These include X-ray crystal structure analysis and the mechanical translation of languages, both of which the authors have helped to develop.

This edition differs little from the first edition (1953), and gives only a very brief account of some of the recent developments in the field, e.g., ferrite core storage and the use of transistors. However, the bibliography, which was a good feature of the first edition, has been considerably expanded to bring it up to date.

R. A. BROOKER.
R. L. GRIMSDALE.

The Mathematics of Diffusion, by J. Crank. Pp. vi + 347. (Oxford: Clarendon Press; London: Cumberlege, 1956.) 50*s*.

The frequency with which diffusion processes are encountered in physics makes the present book by Dr. Crank a valuable addition to the literature of physics. The author sets out to deal with " the description of diffusion processes in terms of solutions of the differential equation for diffusion ". The mechanism of diffusion in particular systems is not discussed nor is the alternative description in terms of ' the random walk '. Throughout his treatment the author had had in mind the practising physicist or technologist who wishes to make use of the results of the mathematical theory.

The book begins with a brief account of the basic hypothesis of the mathematical theory of the diffusion process and with a derivation of the partial differential equation governing such a process in an anisotropic medium.

The next five chapters are concerned with the solution of a number of initial value and boundary value diffusion problems in all of which the diffusion coefficient is assumed to be constant. The main methods available for the solution of the diffusion equation—sources, separation of variables, the Laplace transform—are outlined, with examples, in Chapter II; it is, perhaps, a little suprising to find that the idea of a Green's function is not introduced here and that this concept plays no role in the rest of the book. In Chapters III–VI are derived solutions of the diffusion equation appropriate to a variety of initial and boundary conditions covering diffusion in infinite and semi-infinite media, in plane sheets, in cylinders and spheres. A good variety of special problems is considered and the mathematical solutions derived are well illustrated by diagrams and tables of numerical values.

In Chapter VII the author uses the method developed by Danckwerts to discuss a number of problems in which diffusion occurs in two distinct regions separated by a moving boundary or interface, and in Chapter VIII he considers the equally interesting problems which arise when one substance can diffuse through another with which it reacts chemically.

The next chapter gives an account of some of the methods available for the solution of problems in which the value of the diffusion coefficient varies throughout the medium considered. This leads naturally to the discussion (in Chapter X) of a number of methods which have been developed to provide numerical solutions of diffusion problems (finite difference methods of Schmidt, Crank-Nicolson and Southwell; analogue and other machines).

Chapters XI and XII are given over respectively to the definition and measurement of diffusion coefficients and to some calculated results for variable diffusion coefficients.

The last chapter of the book is an introduction to the theory of the diffusion of a substance through the pores of a solid which absorbs some of it with the evolution (or absorption) of heat.

It will be seen from this resumé of the contents that Dr. Crank has written a book which will be of great interest to physicists and to physical chemists. He must be congratulated in having achieved his aim so successfully—the mathematical treatment is lucid throughout and the physical principles involved in the various diffusion processes are clearly brought out. The problems chosen for discussion seem to be those most likely to be of interest to physicists and their solutions have been evaluated numerically and presented in a graphical form which makes them immediately usable. The only noteworthy omission would seem to be that of any mention of the use in pile theory of solutions of the diffusion equation.

The book is well illustrated and beautifully produced. I. N. SNEDDON.

Théorie Générale de l'Equation de Mathieu et de Quelques Autres Équations différentielles de la Mécanique, by R. CAMPBELL. Pp. xvi+271. (Paris: Masson et Cie, 1955.) Paper bound, 2400 fr., cloth bound, 2900 fr.

The importance of Mathieu's equation in physics and technology has been increasingly recognized in the last two decades. The publication in the Masson collection of mathematical text for physicists of a book on the subject is, therefore, bound to arouse interest, all the more so when it is written by Professor Robert Campbell, who has himself contributed to the theory of such functions.

The book is divided into three parts to which is added a general introduction.

In the introduction the author discusses the occurrence in mathematics of Mathieu's equation.

$$y''+(p-2q\cos 2x)y=0$$

and of the associated equation

$$y''-2y'\tan x=(a=k^2\sin^2 x)y=0.$$

(It is assumed that, in general, x is a real variable.) Applications of these equations in potential theory, elasticity, conduction of heat and electrodynamics are also indicated.

The first part of the book is concerned with the study of the Mathieu functions ce_{2N}, ce_{2N+1}, se_{2N}, se_{2N+1}, of the corresponding functions of Whittaker and Lame-Ince, and of the associated Mathieu functions. In addition to deriving the basic properties of such functions, the author discusses the expansion of them in series of Bessel functions, integral equations, asymptotic expansions for large values of the parameters.

In the second part there is a complete discussion of the Mathieu, Whittaker and MacLauchlan functions of the second kind.

The last part is given over to a very full discussion of the general theory of Mathieu's equation as a special case of Floquet's theory of differential equations with uniform periodic coefficients.

A physicist would find this a difficult book to read but this is due more to the inherent complexities of the subject than to any defect in the author's exposition. Professor Campbell has given an original and full account of the various important aspects of the theory of Mathieu functions, which will be stimulating to pure mathematicians as well as to those who are concerned mainly with applications of these functions to physics and engineering. There is not much overlap with MacLauchlan's well-known treatise. In the restricted field covered by MacLauchlan the present book is not so detailed as MacLauchlan's, though most of the results required in applications are listed; the author's aim has been to set Mathieu's equation in the more general context of the theory of differential equations. This does not make it any the less useful to theoretical physicists but gives it an appeal also to the pure mathematician.

I. N. SNEDDON.

Aberrations of Thin Lenses: An Elementary Treatment for Technicians and Students, by H. H. Emsley. Pp. xvi + 352. (London: Constable, 1956.) 50s.

This book is written for two classes of reader—the technician in industry who has to deal with optical systems and wishes to know the reason for the forms they take, and the student who is not ready for the more advanced books of Conrady and Hopkins, which make a greater demand on his intellectual equipment if not on his mathematical attainments. The book requires a previous knowledge of elementary optics, algebra and trigonometry. The notation is that of Conrady, which has the advantage that it is now well established.

The ground covered comprises the single thin lens, simple eyepieces and the achromatic doublet, treated algebraically, and since the treatment covers 340 pages it will be realised that it is fairly detailed. The book is illustrated by examples worked out and displayed. Trigonometrical computations are also given. There is detailed discussion of the spectacle lens and this is, perhaps, the most valuable part of the book. It must be pointed out however that in the treatment of toric surfaces the equations for refraction on p. 285, and therefore the results, are not correct. A rigid demonstration would not produce a useful result, which is the less to be regretted since, as the author points out, the third order approximations are not sufficiently accurate in any case to give more than very rough ideas.

The diagrams are clear and the expositions generally straightforward and easily understandable, but the use of ' horizontal ' and ' vertical ' to differentiate the meridians is to be deprecated ; ' in and perpendicular to the plane of the paper ' is preferable.

There are very few misprints and those not likely to mislead the student, except for the value of the radian given on p. 45. H. W. L.

CONTENTS OF SECTION A

PAGE

Dr. G. K. Woodgate and Mr. J. S. Martin. Hyperfine Structure in Manganese $^{55}_{25}$Mn 485

Mr. A. M. Arthurs and Mr. J. Hyslop. The Oscillator Strengths of the 1sσ–2pπ, 2pπ–2sσ and 2pπ–3dσ Transitions of HeH^{2+} 489

Dr. I. C. Percival. A Box Variational Method for Scattering Phases . . 494

Dr. A. Dalgarno and Mr. R. McCarroll. Properties of the Hydrogen Molecular Ion: VII—Magnetic Dipole Oscillator Strengths of the 1sσg–3dπg Transition 501

Mr. R. McCarroll and Dr. B. L. Moiseiwitsch. Approximate Molecular Orbitals: VI—The 2pσ state of HeH^{2+} 507

Dr. J. Irving and Mr. F. B. Boucher. A Variation–Iteration Method for Short Range Potential Problems 512

Mr. J. Bowcock. Compound Nuclear Theory and the Optical Model . . . 515

Mr. C. A. ten Seldam. Calculations of the Interatomic Distance at which a Model of Compressed Solid Helium becomes a Metal 529

Research Notes :

Mr. R. D. Amado and Dr. R. J. Blin-Stoyle. Weak Collective Effects in the Nuclear Shell Model 532

Mr. K. R. Allen, Mr. K. Phillips and Miss F. Whittaker. The Scattering of Fast Charged Particles: III—On the Single Scattering of High Energy Positrons and Electrons at Small Angles 536

Prof. D. R. Bates and Miss D. Miskelly. Electron Collision Partial Cross Sections for the 1s–2s and 1s–3s Transitions of Atomic Hydrogen . 539

Prof. J. M. Cassels, Mr. T. W. O'Keeffe, Miss M. Rigby, Mr. A. M. Wetherell and Dr. J. R. Wormald. Experiments with a Polarized Muon Beam 543

Dr. L. Marquez. On the Interpretation of High Energy Fission . . 546

Letters to the Editor :

Dr. P. K. Carroll and Mr. P. J. Daly. Bands Assigned to ZrF . . 549

Mr. D. K. Aitken. A Means of avoiding the ' Electron Rest Mass ' Limitation in a Microton 550

Mr. J. S. Bell. Time Reversal in Beta-decay 552

Corrigendum :

Dr. J. Yoccoz. On the Moments of Inertia of Nuclei 553

Reviews of Books 554

Contents of Section B 556

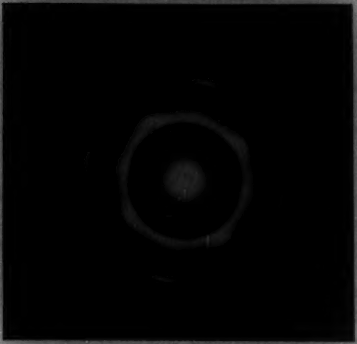


Figure 1. Two-degree {100} orientation of molybdenum (face-centred cubic).

Figure 2. One-degree {100} orientation of molybdenum (face-centred cubic).

Dispersion Effects in Ultrasonic Waveguides and their Importance in the Measurement of Attenuation

By M. REDWOOD

Electrical Engineering Department, Imperial College, London

Communicated by J. Lamb; MS. received 1st *March* 1957, *and in revised form* 29th *March* 1957

Abstract. A study is made of the possible sources of error when the pulse technique is used to determine the absorption of compressional waves (of frequency 10–100 Mc/s) in low-loss solid materials, the specimens being cylindrical and acting as 'waveguides'.

A theoretical analysis is made of a cylindrical guide with transmitting and receiving transducers of arbitrary size: for simplicity the medium is assumed initially to behave as a fluid, and equations describing the possible modes of propagation and their relative amplitudes for a 'piston source' transmitter are developed. The case of transmitting and receiving transducers of diameter equal to that of the guide is treated in detail, and the theoretical voltage at the receiving transducer is plotted as a function of distance from the source; the effects of interference between the several modes of propagation are apparent.

The analysis is extended to solid waveguides, where reflections at the walls of the guide produce a partial conversion to transverse waves. The practical effect of this on mode interference is discussed, and the final results of the analysis are compared with experiment.

The estimated errors in the measurement of attenuation are tabulated for several frequencies and compared with the intrinsic absorption in fused silica and single crystal germanium.

§ 1. INTRODUCTION

THE 'pulse technique' has been used frequently to measure the absorption of ultrasonic waves in solids (Mason and McSkimin 1947, Roderick and Truell 1952, Redwood and Lamb 1956), but only recently has attention been given to the errors in measurement which may be present when investigating materials of very low loss (Seki, Granato and Truell 1956, McSkimin 1956).

The principle of the pulse technique is as follows. A piezo-electric transducer is placed in contact with the specimen, and a 'pulse' consisting of a train of radio-frequency waves of constant amplitude is applied, this train being of a few micro seconds duration and hence containing a large number of cycles of the radio-frequency signal. The transducer converts the electrical signal to a short train of mechanical waves which propagates through the specimen, being continually reflected between the two end faces. When operating as a common transmitter and receiver, the transducer reconverts part of the mechanical energy of successive reflections to electrical signals, which, after amplification and demodulation, are displayed on an oscilloscope (cf. figure 5, Plate). Ideally the amplitude of successive reflections decays exponentially, the rate of decay giving a measure of the energy loss due to absorption in the specimen.

In interpreting results it is generally assumed that propagation takes place at a single frequency only, and continuous wave theory is used in analysis; this is a good approximation for pulses containing a large number of cycles, when the centre portion ('flat top') of the demodulated pulse is used in measurement.

Two experimental methods are of importance. In method (*a*) the dimensions of the specimen in the plane of the transducer are made ideally much greater than those of the transducer, and the boundaries of the specimen parallel to the direction of propagation are assumed to have a negligible effect on the propagating wave, the specimen behaving as if it were unbounded. A loss of energy additional to the absorption in the material occurs due to the 'spreading' of the 'beam' produced by the transducer. In addition the pattern of successive reflections does not show an exponential decay because of diffraction effects. This case has been studied both analytically and experimentally by Seki, Granato and Truell (1956).

In the second method (*b*) used by the author (Redwood and Lamb 1956) the transducer completely covers one end face of the specimen, which is ground in the form of a cylinder of circular cross section; the propagating wave is then 'guided'. Ideally, no energy is lost, other than that due to the intrinsic absorption in the material, but in practice dispersion effects occur due to interference between the several modes which are excited in the specimen, and again the decay pattern is non-exponential.

In the present work the cylindrical 'waveguide' of method (*b*) has been used almost exclusively: it has the disadvantage that the specimen must be accurately ground to have a circular cross section, but the considerable advantages: (i) that small specimens may be used and (ii) that it is possible to use a very large number of reflections (that is, to have a very long effective path length) with consequent improved accuracy in the measured absorption.

It would appear that in method (*a*), in which the transducer is smaller than the specimen, only the first few reflections will behave according to the assumption that propagation takes place as in an unbounded medium, since the beam emitted from the transducer is continually 'spreading'. It would seem advisable, therefore, whatever the size of the specimen, to investigate the effects of the side boundaries; for that reason the general case of a transducer of arbitrary diameter on the end of a cylindrical rod will be analysed here. Extension to the practical case of specimen and transducer of equal diameters is then a simple matter.

For simplicity, in the first part of the analysis, the medium of the guide will be assumed to obey the equations of a perfect fluid, an assumption also made by Seki, Granato and Truell in their analysis of propagation in an unbounded medium. Boundary conditions analogous to those in a solid will be chosen, however; these are that the pressure is zero at the radial boundary.† In addition to the equations governing propagation in a semi-infinite cylindrical guide, the equations for the 'parallel-plate' waveguide will also be discussed. This type of guide has not been used in experimental work, but the equations are closely analogous to those for the cylinder and give a clear physical picture of the 'guiding' of the wave by successive reflections at the bounding walls. The concept thus introduced, that the form of the wave propagating in the guide may be resolved into two plane waves progressing in a zig-zag fashion along the guide by successive reflections at the boundary walls, will be made use of when the important differences between a fluid and a solid are discussed. In the case of a solid a single plane compressional wave incident on a boundary gives rise to *two* reflected waves, one compressional and one transverse.

---

† The practical case of mercury in a steel guide, where the liquid does not wet the walls, obeys these conditions.

§ 2. EQUATIONS GOVERNING THE BEHAVIOUR OF GUIDED WAVES

2.1. *Cylindrical Guide*

2.1.1. *Equations of motion and boundary conditions.*

The differential equation for propagation in a fluid, neglecting viscosity effects may be written (Morse and Feshbach 1953 a)

$$\nabla^2 p = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}$$

where p is the pressure and c is the velocity = [bulk modulus/density]$^{1/2}$.

The required solution of this equation is of the form

$$p(r, z) \propto J_0(\alpha r)\exp(i\gamma z)$$

where

$$\gamma^2 = \frac{\omega^2}{c^2} - \alpha^2 = \left(\frac{\omega}{C_p}\right)^2$$

and C_p is the phase velocity of propagation. Boundary conditions require that at the wall of the cylinder ($r = a_1$) the pressure p shall be zero. For this to apply α must be equal to j_{0n}/a where j_{0n} is any root of the Bessel function $J_0(x) = 0$.

The complete solution may be written

$$p(r, z) = \sum_{n=1}^{\infty} G_n \, J_0\left(j_{0n}\frac{r}{a}\right)\exp(i\gamma_n z)\exp(i\omega t) \qquad \ldots\ldots(1)$$

where G_n is an amplitude constant determined by the pressure distribution produced by the transducer at $z = 0$, and

$$\gamma_n^{\,2} = \left(\frac{\omega}{c}\right)^2 - \left(\frac{j_{0n}}{a}\right)^2 = \left(\frac{\omega}{C_p}\right)^2. \qquad \ldots\ldots(2)$$

Equation (1) represents an infinite number of 'modes' of propagation, each characterized by a pressure distribution across the guide $p(r) \propto J_0(j_{0n}r/a)$ and a phase velocity of propagation

$$C_p = \frac{\omega}{\gamma_n} = c\left[1 - \left(\frac{c}{\omega}\right)^2\left(\frac{j_{0n}}{a}\right)^2\right]^{-1/2}$$

which is always greater than the velocity in an unbounded medium (c).

2.1.2. *Conditions at transmitter and receiver: 'piston' source.*

In order to keep the analysis as general as possible, it will be assumed that two transducers are in use, a transmitter of radius a_1 and a receiver of radius a_2.

(i) *Conditions at the transmitter.*

In order to determine the relative amplitude of each mode (G_n), it is necessary to consider the pressure distribution at $z = 0$. Omitting the term $\exp(i\omega t)$

$$p(r, z=0) = \sum_n G_n J_0\left(j_{0n}\frac{r}{a}\right).$$

To determine G_n both sides of this equation are multiplied by $\mathrm{J}_0(\mathrm{j}_{0m}r/a)r$ and integrated between the limits 0 and a

$$\int_0^a p(r,0)\mathrm{J}_0\left(\mathrm{j}_{0m}\frac{r}{a}\right)r\,dr=\sum_n\int_0^a G_n\mathrm{J}_0\left(\mathrm{j}_{0n}\frac{r}{a}\right)\mathrm{J}_0\left(\mathrm{j}_{0m}\frac{r}{a}\right)r\,dr.$$

The integral on the right hand side of this equation is zero unless $n=m$, when it has the value $\frac{1}{2}a^2\mathrm{J}_1{}^2(\mathrm{j}_{0m})G_n$. Writing $m=n$

$$\int_0^a p(r,0)\mathrm{J}_0\left(\mathrm{j}_{0n}\frac{r}{a}\right)r\,dr=G_n\left(\frac{a^2}{2}\right)\mathrm{J}_1{}^2(\mathrm{j}_{0n}).$$

Hence

$$G_n=\frac{2}{a^2\mathrm{J}_1{}^2(\mathrm{j}_{0n})}\int_0^a p(r,0)\mathrm{J}_0\left(\mathrm{j}_{0n}\frac{r}{a}\right)r\,dr. \qquad\ldots\ldots(3)$$

If the transmitting transducer is centred at $r=0$, and is of radius a_1 ($a_1\leqslant a$), and if it may be assumed that the transducer vibrates in such a manner that the pressure is uniform over its surface†—a condition which, hereafter, will be referred to (loosely) as a 'piston source', then $p(r,0)$ is constant (equal to p_0, say) from $r=0$ to $r=a_1$ and zero from $r=a_1$ to $r=a$, and

$$G_n=\frac{2}{a^2\mathrm{J}_1{}^2(\mathrm{j}_{0n})}\int_0^{a_1}p_0\mathrm{J}_0\left(\mathrm{j}_{0n}\frac{r}{a}\right)r\,dr$$

$$G_n=p_0\,\frac{2a_1}{a\mathrm{J}_1{}^2(\mathrm{j}_{cn})\mathrm{j}_{0n}}\,\mathrm{J}_1\left(\mathrm{j}_{0n}\frac{a_1}{a}\right). \qquad\ldots\ldots(4)$$

(ii) *Conditions at the receiver.*

Substituting for G_n from equation (4) into equation (1) gives the pressure at any point (r,z) in the guide, due to the radiation from a transmitter of radius a_1 at $z=0$. Interest centres on the electrical output from a receiver of radius $a_2(a_2\leqslant a)$, placed in the guide at any value of z. Assuming the electrical output E to be proportional to

$$\int_0^{a_2}p(r,z)2\pi r\,dr,$$

we have

$$E\propto\sum_n 2\pi\int_0^{a_2}G_n\mathrm{J}_c\left(\mathrm{j}_{0n}\frac{r}{a}\right)r\,dr\exp(i\gamma_n z)$$

or

$$E\propto\sum_n\frac{a_1a_2\mathrm{J}_1(\mathrm{j}_{0n}a_1/a)\mathrm{J}_1(\mathrm{j}_{0n}a_2/a)}{\mathrm{j}_{0n}{}^2\mathrm{J}_1{}^2(\mathrm{j}_{0n})}\exp(i\gamma_n z). \qquad\ldots\ldots(5)$$

This equation gives the output voltage from the receiving transducer due to the infinity of modes produced in the guide by the transmitting transducer. Each mode contributes to the resulting voltage E; the contribution is governed by the amplitude factor associated with the mode in question, given by the term

$$\frac{a_1a_2\mathrm{J}_1(\mathrm{j}_{0n}a_1/a)\mathrm{J}_1(\mathrm{j}_{0n}a_2/a)}{\mathrm{j}_{0n}{}^2\mathrm{J}_1{}^2(\mathrm{j}_{0n})},$$

and by the relative phases of the modes with respect to one another, given by the term $\exp(i\gamma_n z)$. For the purpose of computation the output voltage may be represented by an infinite series of vectors (equation (5)), or as many terms of this series as are needed to obtain the required accuracy.

---

† In practice this may not be true, especially at harmonics of the fundamental transducer resonance (Dye 1932).

2.1.3. *Computation of dispersion effects in a waveguide (for $a_1 = a_2 = a$).*

In the case of transducers of radius equal to that of the guide ($a_1 = a_2 = a$) equation (5) reduces to

$$E \propto \sum_n \left(\frac{a^2}{\mathrm{j}_{0n}^2}\right) \exp(i\gamma_n z).$$

(This series has been derived previously by McSkimin (1948).) The relative amplitudes of the first ten terms of this series are given in table 1; the amplitude of the component of E corresponding to the first mode ($n = 1$) is taken as unity.

Table 1. Relative Amplitudes of the Output Voltages associated with the Signals Propagating in the First Ten Modes

n	1	2	3	4	5	6	7	8	9	10
$\dfrac{\mathrm{j}_{01}^2}{\mathrm{j}_{0n}^2}$	1·000	0·189	0·077	0·041	0·025	0·017	0·012	0·009	0·007	0·006

The sum of the series $\Sigma j_{01}^2/j_{0n}^2$ is 1·446, the sum of the first ten terms is 1·388.

It is seen that the amplitude of the output voltage due to the signal propagating in the first mode ($n = 1$, C_p nearest to the velocity in an unbounded medium, c) is 1/1·446 of the total amplitude of the signal voltage at $z = 0$, that is, about 70%. The second mode ($n = 2$) accounts for a further 8%.

In computing the output voltage for various values of z it is necessary to know the angles made by the voltage vectors with some arbitrary reference line. For convenience this may be taken to be the vector representing the first mode ($n = 1$) and the angles for the adjacent modes ($n = 2$ to 10) may be calculated from a simple approximate expression, derived in the following manner

$$\gamma_n = \frac{\omega}{c}\left[1 - \left(\frac{c}{\omega}\right)^2 \left(\frac{\mathrm{j}_{0n}}{a}\right)^2\right]^{1/2} \simeq \frac{\omega}{c}\left[1 - \tfrac{1}{2}\left(\frac{c}{\omega}\right)^2 \left(\frac{\mathrm{j}_{0n}}{a}\right)^2\right] \text{ for large } \frac{a\omega}{c\mathrm{j}_{0n}}.$$

The phase angle between the vector representing mode 1 and the vector representing mode n, at any z is given by

$$(\gamma_1 - \gamma_n)z = \frac{1}{2a^2}\frac{c}{\omega}(\mathrm{j}_{0n}^2 - \mathrm{j}_{01}^2)z. \qquad \ldots\ldots(6)$$

In the present work, a graphical method was chosen as being the most rapid method of computation. The first ten modes were considered, and their behaviour is believed to give a fairly accurate picture of the practical case. For chosen intervals of z, $(\gamma_1 - \gamma_n)z$ was calculated for each mode and the ten vectors drawn at these relevant angles. The length of the resulting vector was measured, the relation between this and the sum of ten terms at $z = 0$ gave the apparent loss of signal due to mode interference. This was converted to decibels and is plotted in figures 1(a) and (b). (In order to keep the applications of this curve as general as possible, the scale chosen was not directly distance from the transmitter z, but $\lambda z/a^2$. Equation (6) may be written

$$(\gamma_1 - \gamma_n)z = \frac{1}{2a^2}\frac{c}{\omega}(\mathrm{j}_{0n}^2 - \mathrm{j}_{01}^2)z = \frac{\lambda}{4\pi a^2}(\mathrm{j}_{0n}^2 - \mathrm{j}_{01}^2)z$$

where λ is the wavelength (at frequency f) in an unbounded medium.

Calculations were made at 100 equally spaced intervals of $\lambda z/a^2$ from $\lambda z/a^2$ to $\lambda z/a^2 = 4{\cdot}3$, (figure 1($a$)), then at a further 90 equal intervals of between $\lambda z/a^2 = 4{\cdot}3$

and 21·5 (figure 1(*b*)). The main interest is in the behaviour near the transducer (figure 1(*a*)), it is hoped that sufficiently close intervals have been considered up to the value $\lambda z/a^2=4{\cdot}3$; beyond that value the curve is intended to give only

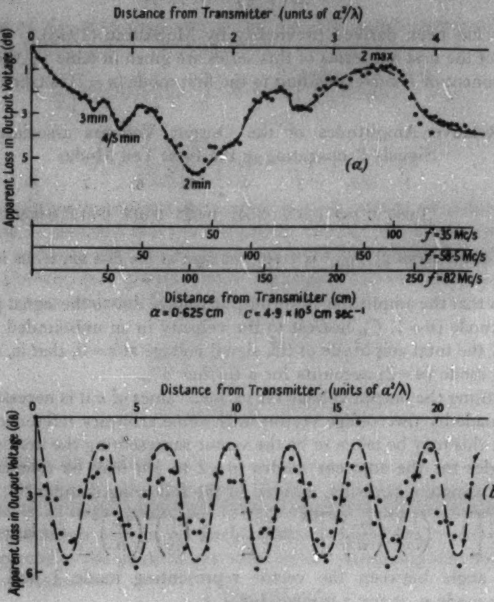


Figure 1.

a general picture (figure 1(*b*)). A further scale has been added to give some idea of practical distances z for values of $c=4{\cdot}9\times10^5$ cm sec^{-1}, and a diameter $2a$ of 12·5 mm, at frequencies of 35, 58·5 and 82 Mc/s, these figures representing the approximate resonance frequencies of a transducer of fundamental frequency 11·7 Mc/s.

The peaks and troughs shown on this plot have a certain physical significance, related to the angles between the vectors representing the voltages associated with different modes. Reinforcement between two modes occurs when the angle between their vectors is a multiple of 2π radians: when it is an odd multiple of π radians destructive interference is greatest. The positions of the trough labelled '2 min' and the peak '2 max' of figure 1(*a*) are determined mainly by the relation between modes 1 and 2 ($n=1$ and $n=2$), which are the predominant modes when a piston source is used. At the point 2 min the angle between the vectors of modes 1 and 2 in π radians, at the point 2 max it is 2π. The general trend of the curve is in fact determined by the behaviour of modes 1 and 2 as can be seen from the form of figure 1(*b*).

The subsidiary peaks and troughs occurring before the main trough 2 min (figure 1(*a*)) are of considerable practical interest as will be seen later; near the point '4/5 min' both fourth and fifth modes are an odd multiple of π radians out of phase with mode 1, point '3 min' corresponds to the third mode being π radians out of phase with mode 1.

Some comment should perhaps be made on the accuracy of this computed curve with its rapid fluctuations. The behaviour is governed almost entirely by the first five modes of the series, the higher modes being of small amplitude and having phase angles (relative to the first mode) which change very rapidly as distance changes, since these angles are proportional to $(\mathrm{j}_{0n}^2-\mathrm{j}_{01}^2)z$. These rapid changes tend to keep the net effect of the higher modes small except possibly at a few 'freak' points, such as the point at $\lambda z/a^2=16$, when all ten vectors were very nearly in phase, giving a resultant almost equal to the sum of ten terms at $z=0$.

It is unlikely that the curve of figure 1 (*a*) and (*b*) could ever be reproduced in detail experimentally, since the assumption that the transmitter behaves as a pure 'piston' source is probably unjustified; deviation from 'piston' behaviour will lead to a change in the relative amplitudes of the modes present, and probably only the larger peaks and troughs would be observed (see also § 4). It should be possible to check the form of the dispersion curve experimentally at least for values of $\lambda z/a^2$ less than about 2, by using a variable length mercury delay line. The results would indicate possibly the validity of the assumption that a vibrating quartz transducer behaves as a 'piston' source.

2.1.4. *Comparison of dispersion and diffraction effects.*

Comparison of the curves computed here for the effects of dispersion in a waveguide with the curve computed by Seki, Granato and Truell for the case of diffraction in an unbounded medium shows a certain similarity in the positions of the first few troughs. In the present work, dispersion minima occur at values of $\lambda z/a^2$ approximately 1·6, 0·85 and 0·6 (though the latter is a trough of small amplitude). In the diffraction case minima are present at $\lambda z/a^2=1{\cdot}6$, 0·9 and 0·6. It is possible to understand the similarity if equation (5) is examined.

Putting $a_1=a_2$

$$E\propto\sum_n\frac{a_1^2}{\mathrm{j}_{0n}^2}\frac{\mathrm{J}_1^2(\mathrm{j}_{0n}a_1/a)}{\mathrm{J}_1^2(\mathrm{j}_{0n})}\exp(i\gamma_n z).$$

For the first few reflections in a short specimen (that is for values of $\lambda z/a^2$ less than about 2, say) an approximation to the 'unbounded' case may be obtained by making a/a_1 large (the guide diameter is $2a$, the transducer diameter is $2a_1$).

The terms of this series that have largest amplitudes will be centred about those with values of n such that $(\mathrm{j}_{0n}a_1/a)=\mathrm{j}_{0m}(m=1, 2, 3\text{ etc.})$. The relative phase angles of these components are given by values of γ_n

$$\gamma_n=\left[\left(\frac{\omega}{c}\right)^2-\left(\frac{\mathrm{j}_{0n}}{a}\right)^2\right]^{1/2}=\left[\left(\frac{\omega}{c}\right)^2-\left(\frac{\mathrm{j}_{0m}}{a_1}\right)^2\right]^{1/2}\quad\begin{array}{l}m=1, 2, 3\text{ etc.}\\ a_1=\text{transducer radius.}\end{array}$$

In comparing diffraction effects with dispersion effects transducers of equal diameters must be assumed; 'a_1' here corresponds, therefore, to 'a' of the previous guide analysis (and of Seki, Granato and Truell's paper).

The angles γ_n between the components of largest amplitude are therefore identical for this case—small diameter transducer, large diameter cylinder—and the dispersion case previously considered (transducer and cylinder of equal diameters): the positions of the initial troughs would be expected to correspond roughly in both cases. The amplitudes of the main components have not been taken into account in this crude comparison and it would not be justified to carry the analysis further.

2.2. *The Parallel Plate Waveguide*

2.2.1. *Equations of propagation.*

The solution in this case is derived in a manner similar to that for a cylindrical guide.

For a parallel plate guide, semi-infinite in the z direction, infinite in the y direction, and with boundaries at $x=0$ and $x=b$ the solution (for $p=0$ at $x=0$ and $x=b$) is (compare with equations (1) and (2))

$$p=\sum_n G_n \sin\left(\frac{n\pi}{b}x\right)\exp(i\gamma_n z)\exp(i\omega t)$$

$$\gamma_n{}^2=\left(\frac{\omega}{c}\right)^2-\left(\frac{n\pi}{b}\right)^2.$$

To obtain G_n, both sides of the equation for p at $z=0$ are multiplied by $\sin(m\pi x/b)$ and integrated between 0 and b:

$$\int_0^b p(x,0)\sin\frac{m\pi x}{b}\,dx=\sum_n\int_0^b G_n \sin\frac{n\pi}{b}x\sin\frac{m\pi x}{b}\,dx.$$

The right-hand side of this equation is zero except when $m=n$, when it is $\frac{1}{2}b$, and so

$$\int_0^b p(x,0)\sin\frac{n\pi x}{b}\,dx=G_n\tfrac{1}{2}b\,.$$

Hence G_n may be determined for any source configuration $p(x, 0)$. For example, for a line source at $x=x_0$ (width dx_0, strength $p(x_0)$)

$$G_n=\frac{2}{b}p(x_0)\sin\frac{n\pi x_0}{b}\,dx_0$$

and

$$p\propto\sum_n p(x_0)dx_0\sin\frac{n\pi x_0}{b}\sin\frac{n\pi x}{b}\exp(i\gamma_n z)\exp i\omega t. \qquad \ldots\ldots(7)$$

It may be of interest to point out that equation (7) may be obtained in a different fashion: propagation in the guide is considered as the sum of a series of reflections of rays generated by a line source situated at $x=x_0$ in the guide. The reflection of these rays at the boundaries is taken into account by placing a double infinity of 'image' sources of appropriate polarity outside the guide. Mathematically the analysis is complex and will not be presented here: the procedure is stated in detail by Morse and Feshbach (1953 b). (The latter have, however, assumed different boundary conditions and their resulting equation 7.2.32 differs slightly from equation (7) of this paper.)

2.2.2. *Reflections at the walls of the guide.*

For any single mode the pressure distribution across the guide is proportional to $\sin(n\pi x/b)$: in the theory of electrical waveguides it is often pointed out that such a distribution may be considered as being produced by two intersecting plane waves, and that a single mode is equivalent to the interference pattern of two plane waves propagating with a velocity c (the velocity in an unbounded medium) and reflected back and forth between the boundaries in a zig-zag path (Lamont 1950).

Consider the wave and reflection shown in figure 2. The pressure for each wave may be represented in the form $A \exp(i\omega t) \exp(i\gamma z')$ where z' is the direction of propagation of the wave considered. The incident wave pressure may be represented by $A \exp\{(i\omega/c)(z \sin\alpha - x\cos\alpha)\} \exp i\omega t$, the reflected wave pressure by $-A \exp\{(i\omega/c)(z\sin\alpha + x\cos\alpha)\} \exp(i\omega t)$ (since at $x=0$ the total pressure must be zero).

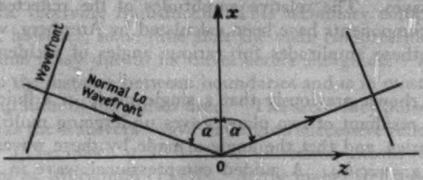


Figure 2.

The resultant becomes $2A \exp\{(i\omega/c)z \sin\alpha\} \sin\{(\omega/c)x\cos\alpha\} \exp(i\omega t)$, that is, across the guide there will be an interference pattern of form $\sin\{(\omega/c)x\cos\alpha\}$. The resultant wave propagates with phase velocity

$$C_p = c/\sin\alpha. \qquad \ldots\ldots(8)$$

To represent the mode 'n', therefore, it is possible to consider two plane waves with angles of incidence α (to the normal) given by

$$\frac{n\pi}{b} = \frac{\omega}{c}\cos\alpha, \quad \text{i.e. } \cos\alpha = \frac{n\pi c}{b\omega}. \qquad \ldots\ldots(9)$$

In practice, for frequencies above about 10 Mc/s, with cylinders of radius of the order of 1 cm, $c/b\omega$ is small, and α is nearly 90° for the important modes, those with values of n below 5. α is closest to 90° for $n=1$. It is also important to note that as ω is increased α approaches 90°.

§ 3. EXTENSION TO GUIDED WAVES IN SOLIDS

3.1. *Boundary Conditions for a Solid*

The most important difference between propagation in a fluid and in a solid is due to the different conditions which hold when reflection takes place at a

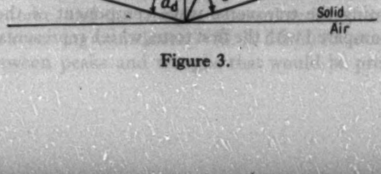


Figure 3.

boundary. In a solid, the reflected wave caused by an incident compressional wave has, in general, two components, a compressional wave and a transverse wave. The angles which the incident and reflected waves make with the normal to the surface are given by the relation (Arenberg 1948) (figure 3)

$$\frac{\sin a_{\mathrm{d}}}{\sin a_{\mathrm{t}}}=\frac{C_{\mathrm{d}}}{C_{\mathrm{t}}} \qquad \ldots\ldots(10)$$

where C_{d} is the velocity of dilational (compressional) waves and C_{t} is the velocity of transverse waves. The relative amplitudes of the reflected compressional and transverse components have been calculated by Arenberg, who gives useful curves relating these amplitudes for various angles of incidence on solid–air boundaries.

It has been shown previously that a single mode in a fluid guide may be regarded as the resultant of two plane waves undergoing multiple reflection at the side boundaries, and that the angle α made by these waves to the normal is given by $\cos\alpha = n\pi c/b\omega$. A guided compressional wave in a solid may be regarded in a similar fashion, putting $\cos a_{\mathrm{d}} = n\pi C_{\mathrm{d}}/b\omega$ but the effect of partial conversion to a transverse wave must be taken into account. The angle a_{t} made by the transverse component is given by $\sin a_{\mathrm{t}} = (C_{\mathrm{t}}/C_{\mathrm{d}})\sin a_{\mathrm{d}}$, and the combination of the transverse waves arising on reflection at both boundaries will produce an interference pattern across the guide, again of the form $\sin\{(\omega/c)x\cos\alpha\}$, but here $C = C_{\mathrm{t}}$, the velocity of transverse waves, and $\cos\alpha = \cos a_{\mathrm{t}}$. The displacement distribution across the guide due to both compressional *and* transverse waves will be of the form

$$A\sin\left(\frac{\omega}{C_{\mathrm{d}}}\cos a_{\mathrm{d}}x\right)+B\sin\left(\frac{\omega}{C_{\mathrm{t}}}\cos a_{\mathrm{t}}x\right). \qquad \ldots\ldots(11)$$

In a previous paper (Redwood and Lamb 1957) the displacement function for a mode propagating in a solid cylindrical guide was developed in a more rigorous fashion, and shown to be of the form

$$A\mathrm{J}_0(\alpha_n r)+B\mathrm{J}_0(\beta_n r);$$

the corresponding equation for a solid parallel plate guide is (Holden 1951)

$$A\sin(\alpha_n x)+B\sin(\beta_n x) \qquad \ldots\ldots(12)$$

where, in both cases,

$$\alpha_n{}^2=\left(\frac{\omega}{C_{\mathrm{d}}}\right)^2\left(1-\frac{C_{\mathrm{d}}{}^2}{C_{\mathrm{p}}{}^2}\right);\quad \beta_n{}^2=\left(\frac{\omega}{C_{\mathrm{t}}}\right)^2\left(1-\frac{C_{\mathrm{t}}{}^2}{C_{\mathrm{p}}{}^2}\right)$$

C_{p} being the phase velocity of propagation.

Substituting for $\cos a_{\mathrm{d}}$ and $\cos a_{\mathrm{t}}$ from equation (8) into equation (11) gives

$$A\sin\left[\frac{\omega}{C_{\mathrm{d}}}x\left(1-\frac{C_{\mathrm{d}}{}^2}{C_{\mathrm{p}}{}^2}\right)^{1/2}\right]+B\sin\left[\frac{\omega}{C_{\mathrm{t}}}x\left(1-\frac{C_{\mathrm{t}}{}^2}{C_{\mathrm{p}}{}^2}\right)^{1/2}\right]$$

which is identical with equation (12).

It was also shown in the previous paper that at frequencies above about 10 Mc/s, for cylinders of radius of the order of 1 cm, the second term of this equation, representing the transverse wave component of the reflecting waves, is generally small compared with the first term, which represents the compressional wave component.

3.2. *The Loss of Energy due to Partial Conversion to Transverse Waves*

In the pulse technique, the conversion of some of the energy of the compressional wave incident on a boundary into a transverse wave has an interesting effect on the decay pattern of successive reflections in a specimen. The angle a_{t} is usually about 30° when a_{d} is near 90°, and the reflected transverse wave travels across the guide with velocity C_{t} (of the order of $\frac{1}{2}C_{\mathrm{d}}$) at this angle to the normal, finally striking the opposite boundary, where it is partially reconverted to a compressional wave which propagates with velocity C_{p} along the guide and is detected by the receiving transducer. This secondary compressional wave generally arrives, however, some time after the main compressional pulse, since it undergoes a time delay due to its travel across the guide. This time delay is proportional to the distance between boundaries and is in most practical cases larger than the duration of the pulses used. The secondary pulse is followed by others generated as the transverse wave undergoes successive reflections between the boundaries. (The effect at 10 Mc/s is illustrated excellently in figures 7 and 8 of the paper by McSkimin (1956)). At the higher frequencies used in the present work, the secondary pulses are scarcely visible, since the angle a_{d} is much nearer 90° and the energy converted to transverse waves is small.

The reflected transverse waves 'drain away' a certain amount of energy from the guided compressional waves, and it is necessary to estimate the loss caused by this effect. Arenberg's curves predict that as frequency increases, the loss due to this conversion will decrease (since $a_{\mathrm{d}} \rightarrow 90°$) and also that the loss from higher order modes ($n > 1$) will be greater than that from the predominant mode ($n = 1$). Since $\cos a_{\mathrm{d}} = n\pi c/b\omega$, a_{d} will be nearer 90° for the first mode than for the higher modes. McSkimin (1956) has developed an approximate method for estimating the loss due to conversion in a cylindrical specimen, but has restricted his analysis to the loss from the first mode. Extending the analysis to cover other modes gives the loss in nepers cm^{-1} as

$$\frac{0{\cdot}545C_{\mathrm{d}}(1-L_n^{2})}{\omega a^{2}}\,\frac{\mathrm{j}_{0n}}{\mathrm{j}_{01}}$$

where L_n is the absolute value of the reflection coefficient for a compressional wave incident at an angle a_{d} and may be obtained from Arenberg's results $\cos a_{\mathrm{d}} = \mathrm{j}_{0n}C_{\mathrm{d}}/a\omega$ in this instance.

Table 2 lists the calculated losses due to partial conversion to transverse waves at the boundaries for each of the first three modes of propagation in a cylinder of fused silica of diameter 1·2 cm. The total loss from a combination of the three modes is also given, for values of z of 0, 100, 200, 500 cm. (The combination of the first three modes chosen is such that their amplitudes are in the ratios that would be produced by a piston source, that is, 1 : 0·189 : 0·077 for modes 1 to 3 respectively.) Four frequencies are considered, 30, 50, 70 and 90 Mc/s. It is immediately apparent that the higher modes are being drained off more rapidly than the first mode and that the signal propagating in the first mode, already predominant at $z = 0$ for a piston source, will be even more predominant after the signal has travelled some way along the solid cylinder. Table 2 also lists the calculated difference in signal amplitude between the peaks where the first two modes reinforce, and the troughs where they interfere, as a function of distance z from the receiving transducer. At 30 Mc/s, for example, the normal 3·3 dB variation between peaks and troughs that would be present if no energy were converted to transverse waves is reduced to 1·2 dB after the signal has propagated through 200 cm of the guide, and to 0·26 dB after 500 cm.

Table 2. Calculated Losses due to Conversion to Transverse Waves at Specimen Walls

	Loss by partial conversion to transverse waves at specimen walls											
1	2			3			4	5			
	$n=1$	2	3	$z=100$	200	500		$z=0$	100	200	500
30	0·0163	0·05	0·100	2·5	4·6	10·0	0·02	3·3	1·94	1·22	0·26
50	0·0073	0·029	0·053	1·2	2·3	5·1	0·01	3·3	2·6	1·94	0·84
70	0·0035	0·016	0·033	0·68	1·14	2·9	0·006	3·3	2·94	2·6	1·66
90	0·0018	0·010	0·022	0·34	0·76	2·0	0·004	3·3	3·1	2·6	1·94

1, Frequency (Mc/s). 2, Mode 'n' (dB cm^{-1}). 3, Combination of the first three modes (dB), assuming 'piston' source, at distances z. 4, Average figure for loss based on loss at 500 cm (dB). 5, Difference in amplitude (dB) between peaks and troughs considering the first two modes ('piston' source), at distance z. Fused silica, diameter 1·2 cm. For a piston source the respective amplitudes of modes $n=1$, 2 and 3 are in the ratio 1 : 0·189 : 0·077.

§ 4. COMPARISON WITH EXPERIMENT

4.1. *Germanium* (*Single Crystal*)

Confirmatory results have been obtained while studying propagation in a cylinder cut from a single crystal of germanium. For the purposes of comparison with theory this crystal must be assumed to behave as an isotropic material: in view of this assumption agreement is fairly close. Very similar results have been obtained in fused silica.

The specimen was 5·52 cm long and of diameter 1·25 cm. The velocity along the axis of the cylinder (the 100 axis of the crystal) was found to be $4{\cdot}91\times10^5$ cm sec^{-1}, and this has been taken as the value of C_d in calculating a^2/λ. For most of the experimental work the transducer covered one end of the specimen completely.

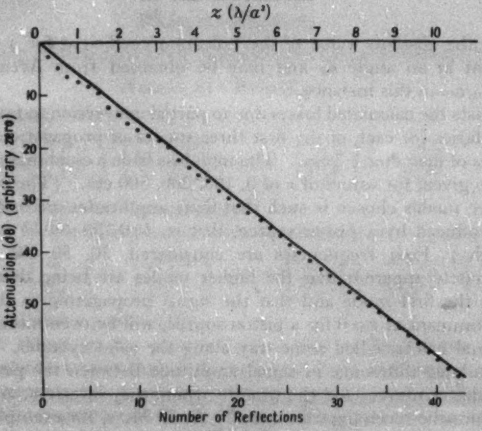


Figure 5.

Figures 4 (*a*)–(*c*) are photographs of the decay patterns of successive reflections at 35, 58·5 and 82 Mc/s, and distinct peaks and troughs are apparent. The distances at which $\lambda z/a^2 = 0{\cdot}9$ and 1·6 (the positions of the two largest troughs predicted theoretically), have been marked on the photographs. It is seen that in each case a distinct trough occurs in the region of both these values of $\lambda z/a^2$. Figure 4(*d*) shows the continuation of the decay at 58·5 Mc/s: it will be seen that it is remarkably smooth after a trough near pulse 9, and a peak near pulse 12 (though the latter is not easily seen).

Figure 5 shows a typical plot of attenuation against distance, obtained at a frequency of 58·5 Mc/s. The accuracy of measurement is governed mainly by a piston attenuator (Redwood and Lamb 1956) and the accuracy of each point relative to the arbitrary zero is estimated at ± 0·1 dB.

The troughs near $\lambda z/a^2 = 0{\cdot}9$ and 1·6 are clearly visible, though in this instance the second trough is broad, and covers $\lambda z/a^2 = 1{\cdot}7$ to 2·2. After a peak at $\lambda z/a^2 = 3{\cdot}5$ there is a slight variation showing some regularity, corresponding to the general variation of figure 1 (*b*), though of considerably less amplitude, about 1 or 2 dB between peaks and troughs compared with the calculated 4 or 5 dB for the fluid guide. It is believed that this is the result of the higher order modes being attenuated to conversion to shear more rapidly than the first mode, as discussed in § 3.2. It is seen that the initial troughs are considerably larger than later troughs, the effects of conversion not being noticeable until the waves have travelled some considerable distance from the source. (The distance to the peak at $\lambda z/a^2 = 3{\cdot}5$ is about 200 cm). In addition the trough at $\lambda z/a^2 = 0{\cdot}9$ is slightly larger in amplitude than theory predicts and the peak at $\lambda z/a^2 = 1{\cdot}1$ is more pronounced, and that at 2·2 is less pronounced. It is believed that these results indicate that the modes are not present here in the proportions computed for a piston source, but that the amplitude of the third, fourth and fifth modes are somewhat larger in the practical case. As has been mentioned previously, it is quite conceivable that the transducer does not vibrate as a 'piston', especially at harmonics of its fundamental resonant frequency, but there is an absence of experimental information concerning this subject in this frequency range.

4.2. *Fused Silica*

In a previous paper (Redwood and Lamb 1956) it was stated that the excellent exponential decays obtained in practice and shown by the remarkably small deviation from linearity in the attenuation plot (as in figure 11 of that paper, and figure 5 of this paper) were believed to be due to the effects of a non-uniform film of oil coupling specimen to transducer. In the present instance, using a germanium specimen it was possible to make an extremely thin coupling film by continual 'wringing' (This was not possible in experiments with fused silica, since continual wringing removed the metal electrode evaporated on to the end face of the silica). The excellent decays obtained in fused silica are now believed to have been obtainable by virtue of the rapid draining away of higher modes in conversion to transverse waves at the radial boundary: the regular deviation from linearity seen in figure 11 of the previous paper is of the order of 2 dB—less than would have been expected theoretically unless some effect were reducing the amplitude of the second mode. The distance between peaks in the results for fused silica also agrees with the computed number of reflections between those peaks that are caused by reinforcement between the first two modes. The initial dispersion peaks and troughs are also present.

4.3. *The Use of a Smaller Receiving Transducer*

In this work a transducer of 1·25 cm diameter was placed on one end of a cylinder of germanium of the same diameter, and used as a transmitter, and a transducer of 6 mm diameter on the other end face used as a receiver. To determine the theoretical position of the peaks and troughs it is necessary to substitute in equation (5)

$$a_1=a \qquad a_2\simeq\tfrac{1}{2}a$$

giving

$$E\propto\sum_n\frac{\mathrm{J}_1(\mathrm{j}_{0n}\tfrac{1}{2})}{\mathrm{j}_{0n}{}^2\mathrm{J}_1(\mathrm{j}_{0n})}\exp(i\gamma_n z).$$

The phase angles between these vectors are the same as for the previous calculations, but the amplitudes of the signal due to different modes has altered. Whereas previously the relative amplitudes were given by the term $1/\mathrm{j}_{0n}{}^2$, they are now given by $\mathrm{J}_1(\tfrac{1}{2}\mathrm{j}_{0n})/\mathrm{j}_{0n}{}^2\mathrm{J}_1(\mathrm{j}_{0n})$. For the case of transmitter and receiver of equal radius the amplitudes of the first four modes were in the ratios $+1:+0{\cdot}189:+0{\cdot}077:+0{\cdot}041$. For the present case, in which the receiver diameter is one half that of the guide and transmitter, the ratios are $+0{\cdot}96:-0{\cdot}241:-0{\cdot}049:+0{\cdot}052$ for modes 1 to 4 respectively. An increase in the effect of mode 2 and a decrease in the effect of mode 3 are apparent. Of most importance, however, is the change in sign of modes 2 and 3, for where reinforcement occurred previously between these modes and mode 1, destructive interference will now be present. The trough near $\lambda z/a^2=1{\cdot}6$ becomes a peak, the peak at $\lambda z/a^2=3{\cdot}2$ a trough.

Figure 4 (*e*) shows the experimental results obtained at 58·5 Mc/s. Here the decay pattern of pulses of larger amplitude are taken from the transmitter crystal and represent the normal behaviour of the system; between these are the pulses of the decay pattern obtained at the small receiving transducer on the other end of the specimen. The photograph shows that the trough near $\lambda z/a^2=1{\cdot}6$ in the normal system (equal size transducers) corresponds to a peak in the case of a receiver of diameter one half that of the guide. A peak is present also at $\lambda z/a^2=0{\cdot}9$, where previously a trough existed. The presence of this peak at $\lambda z/a^2=0{\cdot}9$ tends to suggest that it (and the trough obtained for equal transducers) are due to the effects of the third mode, and that the amplitude of this mode is larger than theory predicts. This hypothesis is supported by the presence of a trough at $\lambda z/a^2=1{\cdot}1$ (the peak normally occurring at this value is produced by mode 3 and mode 1 being in phase).

4.4. *Estimated Errors and Conclusions*

The measured attenuation includes, in addition to the intrinsic absorption in the material (*a*) the loss occurring on reflection at the end faces of the specimen, (*b*) the loss due to the conversion to transverse waves occurring at the side walls of the guide, (*c*) a possible error due to dispersion effects.

(*a*) The loss occurring on reflection at the end faces of the specimen includes energy absorbed by the transducer, coupling film, and the associated electrical circuits, and also a loss due to the partial conversion of the compressional wave to a transverse wave. The latter is extremely small and may be neglected at frequencies above about 10 Ms/c, but absorption of energy in the transducer system can cause an additional attenuation of up to about 1 dB per reflection. (Usually, however, this loss is less than 0·1 dB.) It is possible to evaluate the reflection loss experimentally (Redwood and Lamb 1956). After the initial absorption measurement using a single transducer, a second identical transducer is placed on that end of the specimen which was previously free. Two more measurements of attenuation are then made, transmitting (and receiving) with the second transducer: the first of these two measurements is made with the original transducer still in place, and the second after it has been removed. The difference between these latter two sets of measurements gives the loss on reflection at the original transducer, which is then subtracted from the initial measurement of attenuation to obtain the intrinsic absorption in the material.

(*b*) The main source of error lies in the energy loss due to conversion to transverse waves occurring at the boundary walls of the guide. An estimate of the error has been given in table 2 for a fused silica cylinder of diameter 1·2 cm, but since the loss depends on the diameter of the specimen and Poisson's ratio for the material, it would seem advisable to compute the loss for each individual case. The percentage error depends very much on the intrinsic absorption in the material under test.

The measured attenuations in germanium and fused silica (including conversion to transverse waves) are listed in table 3 for comparison with the figures for the theoretical loss due to conversion, as calculated by the methods of § 3.2. Both these materials have an extremely low intrinsic loss: for the silica the possible error is of the order of 70% at 30 Mc/s, falling to about 2% at 90 Mc/s. For germanium at 30 Mc/s the error will be less (since its intrinsic absorption is higher), possibly between 10 and 20%.

In isotropic materials the loss of energy due to conversion may be calculated, thus giving considerable improvement in the accuracy of the results for intrinsic absorption. Figure 6 shows the attenuation in germanium plotted as a function of frequency, in which corrections have been made to the measured values of absorption, assuming the germanium to behave as an isotropic material. It is seen that the resulting 'intrinsic absorption' curve is reasonably smooth.

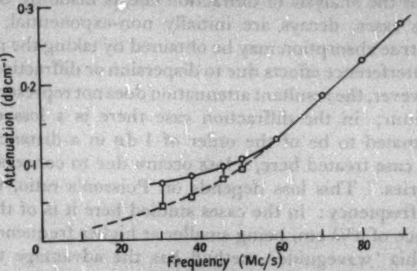


Figure 6.

(*c*) In all the measurements described here, the attenuation has been plotted in decibels for a larger number of reflections in the specimen and the value of absorption used has been that given by a straight line drawn through the peaks of the attenuation plot; this minimizes any error due to dispersion effects.

If only a few reflections are used in the determination of attenuation, however, it is possible to make a considerable error. For example if the reflections considered all lie between $\lambda z/a^2=0$ and 1·5, the slope of the resulting line on the attenuation plot may be in error by approximately 5 dB in the distance z cm, where $\lambda z/a^2=1{\cdot}5$. The estimated possible error due to this is listed in table 3 as a function of frequency; the error may be avoided, of course, by making use of a large number of reflections.

Table 3. Comparison of the Estimated Errors with Measured Absorption for Fused Silica and Germanium

1	2		3	4		5
	Fused SiO_2	Ge		Fused SiO_2	Ge	
30	0·03	0·070	0·02	0·2–1·0	0·05	±0·12
50	0·04	0·095	0·01			±0·09
70	0·08	0·175	0·006			±0·07
90	0·16	0·270	0·004			±0·05

1, Frequency (Mc/s). 2, Measured attenuation (dB cm^{-1}) after correcting for reflection loss, but including loss due to conversion to transverse waves. 3, Estimated average figure for loss due to conversion to transverse waves (dB cm^{-1}). Fused silica, 1·2 cm diameter (table 2). 4, Order of observed reflection loss (dB per reflection); the high reflection loss for silica is due mainly to the resistance of the thin layer of metal which must be evaporated on to one end of the specimen to act as one transducer electrode. This layer is unnecessary when the specimen is a metal. (There is one reflection per double journey in the specimen (10 cm approximately in this instance).) 5, Order of possible error (dB cm^{-1}) due to using an insufficient number of reflections ($a=0{\cdot}6$ cm). The plus sign results if all the reflections considered lie between values of $\lambda z/a^2=0$ and 1·5; the minus sign if only those between $\lambda z/a^2=1{\cdot}5$ and 3 are taken into account.

§ 5. COMPARISON WITH THE EFFECTS OF DIFFRACTION

It is interesting to compare the results of the analysis of dispersion effects with the results of the analysis of diffraction effects made by Seki, Granato and Truell. In both cases, decays are initially non-exponential, but an accurate estimation of the true absorption may be obtained by taking the peaks to represent points at which interference effects due to dispersion or diffraction are a minimum. In both cases, however, the resultant attenuation does not represent the true absorption in the medium; in the diffraction case there is a loss due to the beam 'spreading', estimated to be of the order of 1 dB in a distance of $z=a^2/\lambda$ cm; in the dispersion case treated here, a loss occurs due to conversion to transverse waves at boundaries. This loss depends on Poisson's ratio, the radius of the cylinder, and on frequency: in the cases studied here it is of the order of 0·3 to 0·5 dB in a distance of a^2/λ cm, being smaller at higher frequencies.

In addition this 'waveguide' method has the advantage that a very large number of reflections may be used in the determination of attenuation, and the theoretical analysis is still valid. In the practical cases in which the diffraction theory is employed a specimen diameter of greater than twice that of the transducer is very rarely possible, with the result that the assumption that the medium is 'unbounded' is only valid for points fairly close to the transmitter. Since the guide requires more preparation before it may be used the choice of method depends on the required accuracy.

Where extreme accuracy is required it might be possible to propagate an almost pure single mode by using a shaped electrode with the transmitting transducer to give the required pressure distribution for that mode (Mason and McSkimin 1947). Conversion to transverse waves might be reduced by use of an approximately half-wavelength thickness of a material of a suitable impedance as a coating on the boundary walls of the rod (Isakovitch 1956).

ACKNOWLEDGMENTS

The author wishes to thank Dr. J. Lamb for many helpful suggestions made in the course of this work, and to acknowledge financial assistance from the Worshipful Company of Clothworkers for a research programme of which this work forms a part and from the Oliver Lodge Scholarship of the Institution of Electrical Engineers.

The author is indebted also to Mr. S. E. Bradshaw of the Research Laboratories of the General Electric Company, who was kind enough to provide the single crystal of germanium.

REFERENCES

ARENBERG, D. L., 1948, *J. Acoust. Soc. Amer.*, **20**, 1.
DYE, W. D., 1932, *Proc. Roy. Soc.* A, **138**, 1.
HOLDEN, A. N., 1951, *Bell Syst. Tech. J.*, **30**, 956.
ISAKOVITCH, M. A., 1956, *Akust. Zh.*, **2**, 150.
LAMONT, H. R. L., 1950, *Waveguides*, 3rd Edn. (London : Methuen), pp. 68–70.
MASON, W. P., and MCSKIMIN, H. J., 1947, *J. Acoust. Soc. Amer.*, **19**, 464.
MCSKIMIN, H. J., 1948, *J. Acoust. Soc. Amer.*, **20**, 418 ; 1956, *Ibid.*, **28**, 484.
MORSE, P. M., and FESHBACH, H., 1953 a, *Methods of Theoretical Physics* (New York : McGraw Hill), p. 163 ; 1953 b, *Ibid.*, pp. 812–820.
REDWOOD, M., and LAMB, J., 1956, *Proc. Instn Elect. Engrs*, Pt. B, **103**, 773; 1957, *Proc. Phys. Soc.* B, **70**, 136.
RODERICK, R. L., and TRUELL, R., 1952, *J. Appl. Phys.*, **23**, 267.
SEKI, H., GRANATO, A., and TRUELL, R., 1956, *J. Acoust. Soc. Amer.*, **28**, 230.

The Atomic Heats of Calcium, Strontium and Barium between 1·5° and 20°K

By L. M. ROBERTS
Royal Radar Establishment, Malvern, Worcs.

MS. received 23rd April 1957

Abstract. The atomic heats of calcium, strontium and barium have been measured in the temperature range 1·5° to 20°K, with the primary object of obtaining values for the electronic contributions to the specific heats. In all three metals at the lowest temperatures an atomic heat of the form $C = AT + 1943(T/\Theta)^3$ joule deg^{-1} (g atom)$^{-1}$ is found, where A and Θ are respectively $(2{\cdot}73 \pm 0{\cdot}06) \times 10^{-3}$ and $228{\cdot}9 \pm 1{\cdot}7$°K for calcium, $(3{\cdot}64 \pm 0{\cdot}18) \times 10^{-3}$ and $147{\cdot}0 \pm 1{\cdot}2$°K for strontium and $(2{\cdot}7 \pm 0{\cdot}5) \times 10^{-3}$ and $110{\cdot}5 \pm 1{\cdot}8$°K for barium. The experimental values of A, which represent the electronic contribution to the atomic heat, are compared with those calculated on the basis of the Sommerfeld free-electron model, and are found to be between 1·4 and 2·0 times as great, indicating a considerable overlapping of the first two Brillouin zones.

Smoothed values of the atomic heats are tabulated at intervals up to 20°K and the variations of Θ with temperature are shown graphically. The internal energies, entropies and free energies are also tabulated.

§ 1. Introduction

In atoms of the alkaline earth elements the electrons are arranged in closed shells with two outer valence electrons which, in the solid state, might be expected to fill completely the first Brillouin zone. In this event the alkaline earths would be electrical insulators. The fact that they are metals indicates that the first Brillouin zone must overlap the next, and therefore the way in which the density of states varies with energy will be complicated. If the overlap is small, the density of states at the Fermi surface will be small also.

Some information on this should be provided by measurements of the electronic specific heat, which is given, per unit volume, by

$$c = \frac{2\boldsymbol{k}^2}{3} \pi^2 T n(E) \qquad \ldots\ldots(1)$$

where $\boldsymbol{k}$ is Boltzmann's constant and $n(E)$ the density of states at the Fermi surface per unit volume of the metal (Wilson 1936). At normal temperatures this electronic specific heat is very small in comparison with the heat capacity of the crystal lattice; but at low temperatures, the lattice specific heat falls rapidly, and ultimately becomes proportional to the cube of the absolute temperature. The total atomic heat can then be written

$$C = 1943(T/\Theta)^3 + AT \text{ joule deg}^{-1} \text{ (g atom)}^{-1} \qquad \ldots\ldots(2)$$

where $A = 6{\cdot}265 \times 10^{-46} V n(E)$, Θ is the Debye characteristic temperature of the lattice and V the atomic volume, so that at sufficiently low temperatures the electronic term AT will predominate. Thus, measurement of the atomic heat as a function of temperature in this range leads to values of Θ and $n(E)$, and hence to an estimate of the degree of overlap of the Brillouin zones.

The atomic heats of calcium, strontium and barium have therefore been measured between 1·5° and 20°K.

§ 2. Experimental Details

The specimens were supplied by Messrs. New Metals and Chemicals Ltd., and spectrographic analysis by Messrs. Johnson, Matthey and Co. showed the following impurities: in the calcium, Ba 0·1%, Al, Fe, Mg, Mn, Si, Sn, Sr 0·05% each; in the strontium, Ba, Fe, Mn 0·5% each, Ca 0·2%, Si 0·05%; and in the barium, Sr 0·2%, Fe 0·05%, Sn, Al 0·02% each. The calcium sample was a roughly rectangular block 4·5 × 3 × 2 cm weighing 43·8 g (1·09 g atom). The strontium and barium samples each consisted of two cylinders approximately 2 cm diameter by 3 cm long, the total weight of the strontium being 60·2 g (0·687 g atom) and of barium, 80·5 g (0·586 g atom). Having cut a sample to size, most of the surface oxide was removed in the air, or, in the case of barium, in a box containing solid carbon dioxide. The remaining oxide was scraped off, and the sample was placed in the calorimeter in a handling box through which pure helium gas was circulating. The calorimeter lid was then soldered on with Wood's metal and, after reducing the pressure of helium inside to approximately 0·2 mm of mercury, the calorimeter was finally sealed off. After completing the measurements the calorimeter was reopened in the handling box and the specimen was transferred to a weighing bottle.

The apparatus and general method have already been described (Parkinson 1954) and details of the calorimeter, which was made of copper and was fitted with a carbon thermometer, have also been given (Parkinson and Roberts 1957). The heat capacity of the calorimeter was known to within 5% from separate measurements and was at no time greater than 20% of the total measured heat capacity. A small correction was applied for the variable amount of Wood's metal used in sealing the calorimeter lid. The heat capacity measurements themselves are believed to be accurate to 1%, and the total possible error in the final values is therefore about 2%.

§ 3. Results

Monotonically increasing atomic heats were obtained for all three metals, and smoothed values are given in table 1.

According to equation (2), a plot of C/T against T^2 at sufficiently low temperatures should give a straight line of which the slope is a measure of the lattice contribution, and the intercept on the C/T axis a measure of the electronic contribution to the specific heat. Such a plot for calcium is shown in figure 1, and a linear relationship is found below 4·9°K. The increased scatter in the points below 2·4°K was caused by desorption from the surfaces of the specimen and calorimeter of the helium gas used for thermal exchange between them. The straight line shown was obtained by a least squares analysis applied to points between 2·4° and 4·9°K, and corresponds to an atomic heat of the form of equation (2) with $\Theta = 228{\cdot}9$°K and $A = 2{\cdot}73 \times 10^{-3}$ joule deg^{-2} (g atom)$^{-1}$.

For strontium (figure 2), linear behaviour was found below 4·3°K, and the straight line shown, obtained exactly as for calcium, corresponds to $\Theta = 147{\cdot}0$°K and $A = 3{\cdot}64 \times 10^{-3}$ joule deg^{-2} (g atom)$^{-1}$.

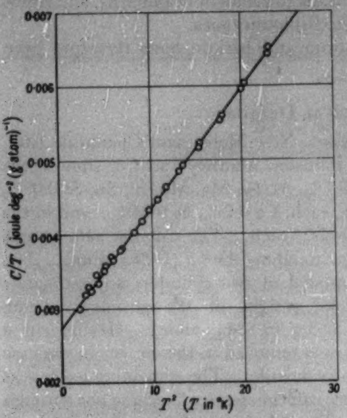


Figure 1. Calcium. The straight line corresponds to $C=(2{\cdot}73\times10^{-3})\,T+1943(T/228{\cdot}9)^3$.

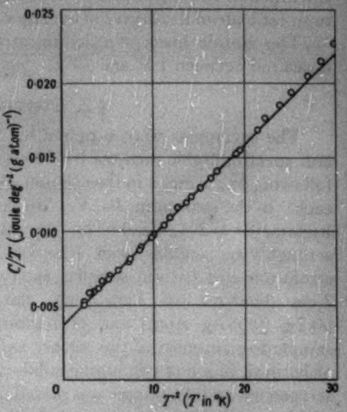


Figure 2. Strontium. The straight line corresponds to $C=(3{\cdot}64\times10^{-3})\,T+1943(T/147{\cdot}0)^3$.

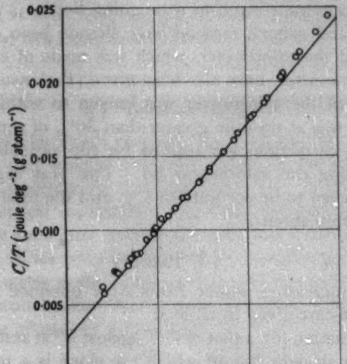


Figure 3. Barium. The straight line corresponds to $C=(2{\cdot}7\times10^{-3})T+1943(T/110{\cdot}5)^3$.

Similarly, for barium (figure 3), using points between 2·4° and 3·0°K, the atomic heat is given by $\Theta=110{\cdot}5$°K and $A=2{\cdot}7\times10^{-3}$ joule deg^{-2} (g atom)$^{-1}$.

If points below 2·4°K are included in the least squares analysis, the values obtained for Θ and A become respectively 229·8°K and $2{\cdot}76\times10^{-3}$ joule deg^{-2} (g atom)$^{-1}$ for calcium, 148·3°K and $3{\cdot}85\times10^{-3}$ joule deg^{-2} (g atom)$^{-1}$ for strontium, and 112·3°K and $3{\cdot}3\times10^{-3}$ joule deg^{-2} (g atom)$^{-1}$ for barium.

Table 1. The Smoothed Atomic Heats of Calcium, Strontium and Barium

T		C_p		T		C_p	
	Ca	Sr	Ba		Ca	Sr	Ba
1·5	0·00464	0·00752	0·0090	7	0·081	0·253	0·665
2·0	0·00676	0·0122	0·0169	8	0·118	0·385	1·00
2·5	0·00935	0·0186	0·0293	9	0·158	0·544	1·38
3·0	0·0126	0·0274	0·0471	10	0·205	0·74	1·85
3·5	0·0165	0·0389	0·0725	12	0·348	1·27	2·94
4·0	0·0213	0·0537	0·107	14	0·550	1·99	4·11
4·5	0·0270	0·072	0·154	16	0·833	2·81	5·44
5·0	0·0344	0·096	0·220	18	1·19	3·76	6·80
6·0	0·0540	0·161	0·408	20	1·63	4·79	8·16

C_p in joule deg^{-1} (g atom)$^{-1}$, T in °K

§ 4. DISCUSSION

The primary aim of the present measurements was to obtain values for the electronic contributions to the specific heats by extrapolating to zero temperature the linear portions of graphs of C/T against T^2. This method is most accurate for metals with a high Debye temperature, for then, not only is the lattice contribution smaller, but also the graph is likely to be linear to a higher temperature.

Owing to smoothing of the resistance thermometer calibration, the random errors in the measured atomic heats are small and most of the estimated error of 2% is systematic. The straight lines in figures 1, 2 and 3 were drawn using points from 2·4°K up to the limit of the linear range. If the systematic error is a simple bias of, say, 2% on all the points, the intercept will also be wrong by 2%, but if, unfortunately, the error were such as to bias the points at either end in opposite directions, the line would be tilted and the error in the ordinate intercept could be appreciably increased, particularly for barium where the linear range is small. The different intercept obtained by including the less reliable points below 2·4°K (§ 3) also provides a check on the probable accuracy of the experimental values, and it so happens that this difference, for strontium and barium, is about equal to the possible 'tilt' error. For calcium however, the simple bias has the largest effect. It is on these considerations that the following electronic specific heats have been estimated: calcium $(2{\cdot}73 \pm 0{\cdot}06)T \times 10^{-3}$ joule deg^{-1} (g atom)$^{-1}$; strontium $(3{\cdot}64 \pm 0{\cdot}18)T \times 10^{-3}$ joule deg^{-1} (g atom)$^{-1}$; barium $(2{\cdot}7 \pm 0{\cdot}5)T \times 10^{-3}$ joule deg^{-1} (g atom)$^{-1}$.

Table 2

	$n(E)_{\text{exp}}$	$n(E)_{\text{Som}}$	$n(E)_{\text{exp}}/n(E)_{\text{Som}}$
Calcium	$16{\cdot}8 \times 10^{40}$	$9{\cdot}2 \times 10^{40}$	1·82
Strontium	16·8	8·4	2·01
Barium	11·8	8·2	1·44

$n(E)$ in joule^{-1} cm^{-3}.

By comparing these electronic specific heats with the linear term of equation (2) it is possible to obtain experimental values for the density of states at the Fermi surface ($n(E)_{\text{exp}}$), and these are given in table 2. Also given in the table for comparison are values ($n(E)_{\text{Som}}$) calculated from the expression for $n(E)$ appropriate to the Sommerfeld free-electron model, that is (Wilson 1936),

$$n(E) = \frac{2\pi m}{h^2}\left(\frac{3p}{\pi}\right)^{1/3}$$

where m is the mass of an electron, h Planck's constant, and p the number of electrons per unit volume, two electrons per atom being assumed.

The ratio $n(E)_{\exp}/n(E)_{Som}$ is considerably greater than unity for all three metals. Hence, since the density of states in the first Brillouin zone must fall below the free-electron value towards the top of the zone, it appears that the second Brillouin zone is making a considerable contribution to the total density of states. The calculations of Manning and Krutter (1937) for calcium show a very steep rise in the density of states at the bottom of the second zone, but the degree of overlap indicated in their paper is insufficient to account for our observed values. Because of this steep rise however, a small alteration in the overlap of the two zones would affect the density of states at the Fermi surface very appreciably, and these authors point out that the particular overlap obtained may depend rather critically on the approximations used in the calculations.

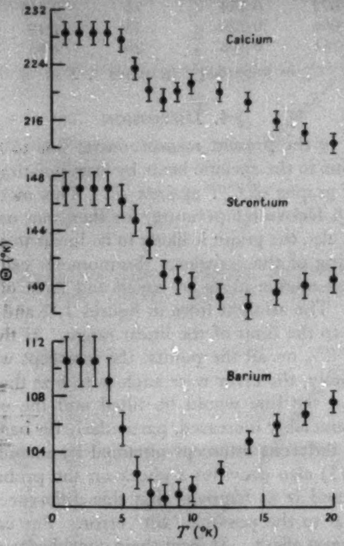


Figure 4. Debye temperatures. The vertical lines represent the estimated uncertainties.

Similar arguments presumably apply to strontium and barium.

The lattice contributions to the atomic heats, in terms of their equivalent Debye temperatures, are shown in figure 4. These temperatures have been calculated after subtracting the above electronic contributions from the smoothed atomic heats of table 1. Uncertainties in the limiting values of Θ as T tends to 0°K have been estimated on the same basis as those in the electronic specific heats, giving $228\cdot9 \pm 1\cdot7$°K for calcium, $147\cdot0 \pm 1\cdot2$°K for strontium and $110\cdot5 \pm 1\cdot8$°K for barium. Towards 20°K the uncertainties decrease as the electronic term becomes less important. This is shown in figure 4. The variation of Θ with temperature is of the usual form, showing an initial decrease as the temperature rises (Blackman 1955). However, a detailed comparison of the shapes of the curves is not profitable owing to the differing crystal structures of the three metals.

The only previous measurements on the specific heats of the alkaline earth metals below 20°K are those of Clusius and Vaughan (1930) on calcium from 10°K upwards. The present results are between 6% and 10% higher than theirs and appear to be rather more consistent with the measurements of Günther (1915), which start at 22°K.

Table 3. The Internal Energy, Entropy and Free Energy of Calcium, Strontium and Barium

	Calcium			Strontium			Barium		
T	U	S	$-F$	U	S	$-F$	U	S	$-F$
1	0·0014	0·0028	0·0014	0·0020	0·0038	0·0019	0·0017	0·0032	0·0015
2	0·0061	0·0059	0·0057	0·0097	0·0089	0·0081	0·0112	0·0093	0·0074
3	0·0156	0·0096	0·0134	0·0287	0·0164	0·0205	0·0414	0·0211	0·0220
4	0·0322	0·0144	0·0253	0·068	0·0276	0·0422	0·115	0·0423	0·0536
5	0·0594	0·0204	0·0426	0·141	0·0438	0·078	0·234	0·077	0·113
6	0·103	0·0284	0·067	0·268	0·067	0·133	0·581	0·133	0·219
7	0·170	0·0386	0·101	0·473	0·098	0·216	1·11	0·215	0·393
8	0·269	0·0518	0·146	0·79	0·141	0·336	1·94	0·325	0·66
9	0·403	0·0676	0·205	1·25	0·195	0·503	3·12	0·464	1·06
10	0·584	0·087	0·282	1·89	0·262	0·732	4·73	0·634	1·61
12	1·13	0·136	0·503	3·87	0·441	1·42	9·5	1·07	3·29
14	2·01	0·204	0·84	7·1	0·690	2·54	16·6	1·61	5·96
16	3·38	0·294	1·33	11·9	1·01	4·24	26·1	2·24	9·8
18	5·40	0·413	2·04	18·5	1·39	6·61	38·3	2·96	15·0
20	8·23	0·562	3·01	27·0	1·84	9·9	53·3	3·75	21·7

Note : The heat units are joule (g atom)$^{-1}$, and the temperatures are in degrees Kelvin.

ACKNOWLEDGMENTS

The author wishes to thank Dr. D. H. Parkinson for suggesting this work and for his continued interest in it. She also wishes to thank Mr. D. C. Handley and Mr. P. E. Hartwright for experimental assistance and Mrs. D. G. Maisey for programming the electronic computer for the least squares analyses. This paper is published by permission of the Controller, H.M. Stationery Office.

REFERENCES

BLACKMAN, M., 1955, *Handbuch der Physik*, **7**, 1, 361-372 (Berlin : Springer).
CLUSIUS, K., and VAUGHAN, J. V., 1930, *J. Amer. Chem. Soc.*, **52**, 4686.
GÜNTHER, P., 1915, *Ann. Phys., Lpz.*, **51**, 828.
MANNING, M. F., and KRUTTER, H. M., 1937, *Phys. Rev.*, **51**, 761.
PARKINSON, D. H., 1954, *J. Sci. Instrum.*, **31**, 178.
PARKINSON, D. H., and ROBERTS, L. M., 1957, *Proc. Phys. Soc.* B, **70**, 471.
WILSON, A. H., 1936, *The Theory of Metals* (Cambridge : University Press), p. 185.

The Atomic Heats of Lithium, Sodium and Potassium between 1·5° and 20°K

By L. M. ROBERTS
Royal Radar Establishment, Great Malvern, Worcs.

MS. received 1*st March* 1957

Abstract. The atomic heats of lithium, sodium and potassium have been measured between 1·5° and 20°K, chiefly with the object of determining the effective mass of the conduction electrons. For lithium it was found that, below 4°K, the atomic heat could be written as $C=(0{\cdot}418T+0{\cdot}0092T^3)\times10^{-3}$ cal mol^{-1} deg^{-1}, corresponding to an effective mass m^* for the conduction electrons of 2·32 times the free electron mass m_0, and a Debye characteristic temperature Θ for the lattice of 369°K. Near 15°K a small anomaly in the atomic heat was observed. For sodium below 4·5°K the atomic heat was of the form

$$C=(0{\cdot}328T+0{\cdot}117T^3)\times10^{-3}\ \text{cal mol}^{-1}\ \text{deg}^{-1},$$

corresponding to $m^*=1{\cdot}22m_0$ and $\Theta=158$°K, and no anomalies were found. For potassium it was not possible to analyse the atomic heat so exactly, but a value of m^* of about $1{\cdot}1m_0$ is estimated, with $\Theta=89$°K.

Smoothed values of the atomic heats are tabulated at intervals up to 20°K, and calculated values of Θ are given graphically. The internal energy, entropy and free energy are also tabulated.

§ 1. Introduction

The alkali metals are interesting because of their simple electronic structure, each nucleus being surrounded by a very stable electron configuration plus only one outer valence electron, which in the solid becomes free to move through the lattice. In these circumstances the conduction band is only half full, and its electrons might be expected to approximate more closely than in more complex metals to the free electron gas of the simple Sommerfeld model.

The contribution of the electrons to the specific heat per unit volume of a metal is given by

$$c_V=\tfrac{1}{3}\pi^2k^2Tn(E) \qquad \ldots\ldots(1)$$

where $n(E)$ is the density of states per unit volume at the Fermi surface (Wilson 1936). Hence the measurement of electronic specific heat gives a direct measure of the density of states. For the free-electron model, $n(E)$ is given by

$$n(E)=\frac{2\pi m}{h^2}\left(\frac{3p}{\pi}\right)^{1/3} \qquad \ldots\ldots(2)$$

where m is the mass of an electron, and p is the number of electrons per unit volume. Any difference between the observed value and that given by (2), assuming for the alkali metals one electron per atom, is a measure of the deviation from the simple model. Such deviations may be represented by associating a different mass m^*, the 'effective mass', with the electrons in the metal.

In order to obtain an experimental value for c_V it is necessary to be able to separate it from the specific heat of the lattice which at normal temperatures is by far the larger quantity. At very low temperatures, however, the lattice contribution falls rapidly and ultimately should become proportional to the cube of the absolute temperature, so that the total atomic heat can be written as

$$C = AT + 464{\cdot}5(T/\Theta)^3 \text{ cal mol}^{-1} \text{ deg}^{-1} \qquad \ldots\ldots(3)$$

where Θ is the Debye temperature of the lattice,

$$A = 3{\cdot}25 \times 10^{-5} V^{2/3} m^*/m_0 \text{ cal mol}^{-1} \text{ deg}^{-2}$$

and V is the atomic volume in cm^3 mol^{-1}. Experimental measurement of the atomic heat should therefore enable values of Θ and A, and hence of m^*, to be determined.

Such a method has been used recently by Parkinson and Quarrington (1955) to obtain a value for the effective mass of the electrons in sodium, and the purpose of the present work was to repeat their measurements with higher precision and to extend them to other alkali metals.

§ 2. METHOD

A description of the apparatus and general method has been given previously (Parkinson 1954), and the calorimeter has also been described (Parkinson and Roberts 1957). The temperature of the calorimeter and specimen was measured with a carbon thermometer, and was raised in steps varying from 0·2° at 1·3°K to 1° at 20°K, whilst the shield temperature was held constant at 2°K, 4·2°K or 20°K as convenient. Corrections were made for residual heat leaks by observing the temperature drift between heating periods. The heat capacity of the calorimeter, containing only exchange gas, was measured with an accuracy to about 2% in a separate series of measurements, and the heat capacities of the pure nickel containers in which the specimens were cast (see below) were calculated from published data (Keesom and Clark 1935, Clusius and Goldman 1936). A small correction was also made for the varying amount of Wood's metal used in sealing on the calorimeter lid. The smoothed atomic heats given in table 1 are estimated to be correct to 1% except below 2·5°K.

§ 3. PREPARATION OF SAMPLES

A block of lithium was supplied, sealed under argon, by Messrs Johnson, Matthey & Co. Ltd. The chief impurities were 0·3%–0·5% sodium and about 0·05% potassium. The sodium and potassium samples, stated purity 99·995%, were supplied by the Atomic Energy Research Establishment, Harwell, in sealed glass capsules into which they had been distilled. In each instance the container was opened in a handling box, through which a stream of pure helium flowed continuously, and the metal was melted in an iron crucible and poured into a thin-walled weighed nickel container in which it solidified. This was then placed inside the calorimeter and the lid soldered on with Wood's metal. After the pressure of helium inside it had been reduced to 2 mm Hg the calorimeter was finally sealed off and removed from the box. Even using the handling box, a little oxidation occurred during the melting process. With sodium, however, the bright metal poured out of the crucible leaving the oxide scum behind, but with the lithium and potassium it was impossible to avoid pouring a little of the oxide into the container with the metal.

After completing the heat capacity measurements, the calorimeter was reopened inside the handling box and the metal specimen in its nickel container was transferred to a weighing bottle. The weights of the specimens and their containers respectively were: lithium 10·16 g (1·46 mol), 1·03 g; sodium 36·31 g (1·58 mol), 1·07 g; potassium 21·07 g (0·539 mol), 0·86 g. The same calorimeter was used for all three specimens.

§ 4. Results

(i) *Lithium*

Five series of measurements were made on the lithium sample, during which each part of the temperature range was covered at least three times. Reproducible results were obtained between 2·4° and 13°K, and above 17°K.

At sufficiently low temperatures according to (3) a plot of C/T against T^2 should give a straight line, the intercept on the C/T axis being a measure of the electronic contribution to the atomic heat, and the slope a measure of the lattice contribution. Such a plot for lithium is shown in figure 1 and linear behaviour is found between 2·4° and 4·5°K. A least squares method of analysis applied to these points gives the straight line shown which corresponds to an atomic heat of the form

$$C=0{\cdot}000418T+464{\cdot}5(T/369)^3 \text{ cal mol}^{-1}\text{ deg}^{-1}. \qquad \ldots\ldots(4)$$

Below 2·4°K the measured values tend to be high and rather scattered. This is believed to be a spurious effect caused by adsorbed helium (see § 4 (iv)). Inclusion of these points in the least squares analysis gives, in place of (4),

$$C=0{\cdot}000435T+464{\cdot}5(T/386)^3 \text{ cal mol}^{-1}\text{ deg}^{-1}. \qquad \ldots\ldots(4a)$$

The irreproducibility between 13° and 17°K is shown in figure 2. For the points of series 1, 2 and 3 the normal procedure was followed of cooling the calorimeter to 77°K the previous evening. The points so obtained are in general high, and after each heating interval a time of up to 2 minutes was required to reach thermal equilibrium. For series 4 and 5 the apparatus was cooled from room temperature on the morning of the experiment and the calorimeter was cooled as rapidly as possible (40 minutes) from room temperature to 20°K. The points thus obtained lie more or less on a smooth curve joining the upper and lower reproducible atomic heat values, and there was no noticeable time lag in reaching thermal equilibrium after heating.

Smoothed values of the measured atomic heat are given in table 1, and values of the Debye characteristic temperature Θ, calculated after subtraction of the electronic specific heat from (4), are shown in figure 5.

Table 1

T(°K)	C_p (mcal mol^{-1} deg^{-1})			T(°K)	C_p (cal mol^{-1} deg^{-1})		
	Li	Na	K		Li	Na	K
1·5	0·66	0·89	3·03	8	0·0091	0·067	0·389
2·0	0·91	1·59	6·3	9	0·0119	0·097	0·520
2·5	1·27	2·65	12·0	10	0·0152	0·131	0·660
3·0	1·50	4·15	20·7	12	0·0238	0·218	0·99
3·5	1·86	6·17	33·8	14	—	0·347	1·32
4·0	2·27	8·8	51·3	16	—	0·508	1·65
5·0	3·35	16·4	103	18	0·0722	0·679	2·00
6·0	4·91	27·9	179	20	0·0953	0·850	2·36
7·0	6·84	44·8	276				

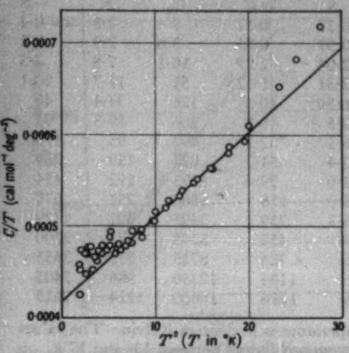


Figure 1. Lithium. The full line corresponds to $C = 0{\cdot}000418T + 464{\cdot}5(T/369)^3$ cal mol^{-1} deg^{-1}.

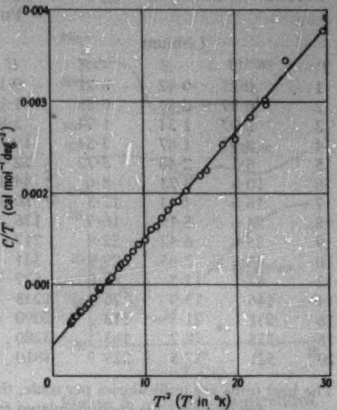


Figure 3. Sodium. The full line corresponds to $C = 0{\cdot}000328T + 464{\cdot}5(T/158{\cdot}2)^3$ cal mol^{-1} deg^{-1}.

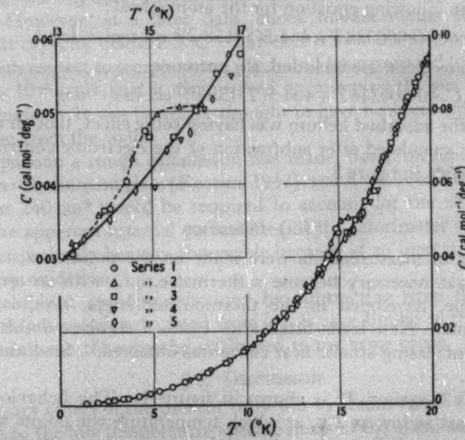


Figure 2. The atomic heat of lithium. Below 10°K several points have been omitted for clarity.

(ii) *Sodium*

Four series of measurements were made on the sodium sample, the first and last covering the whole range from 1·3° to 20°K. The atomic heat increased monotonically with temperature, and smoothed values are given in table 1.

Table 2. Internal Energy, Entropy and Free Energy of Lithium, Sodium and Potassium

	Lithium			Sodium			Potassium		
T	U	S	$-F$	U	S	$-F$	U	S	$-F$
1	0·21	0·42	0·21	0·19	0·37	0·17	0·4	0·8	0·4
2	0·87	0·86	0·85	1·12	0·97	0·81	3·6	2·9	2·2
3	2·07	1·34	1·94	3·84	2·04	2·27	16·1	7·8	7·3
4	3·93	1·87	3·54	10·1	3·81	5·12	51	17·7	20·1
5	6·7	2·48	5·72	22·4	6·54	10·3	126	34·4	46
6	10·8	3·23	8·6	44·3	10·5	18·8	266	59·8	93
7	16·6	4·13	12·3	80·3	16·1	32·1	492	95	170
8	24·6	5·19	16·9	136	23·4	51·3	823	139	289
9	35·0	6·42	22·8	217	33·0	79·4	1275	192	454
10	48·6	7·85	29·9	331	45·0	118	1860	254	679
12	87·1	11·3	49·0	677	76·3	239	3495	402	1329
14	146	15·9	76·1	1235	119	432	5775	577	2300
16	231	21·5	113	2090	176	725	8715	773	3655
18	353	28·7	163	3280	246	1144	12350	986	5405
20	521	37·5	229	4810	326	1714	16650	1214	7615

The heat units are millicalories per mole, the temperature is in degrees Kelvin. The values for lithium have been calculated from the smooth curve between 13° and 17°K.

A plot of C/T against T^2 is shown in figure 3. Any effect due to adsorbed helium is masked here by the higher lattice specific heat of sodium. Using points corresponding to temperatures between 2·5° and 4°K, a least squares analysis gives the following equation for the atomic heat:

$$C = 0{\cdot}000328T + 464{\cdot}5(T/158{\cdot}2)^3 \text{ cal mol}^{-1}\text{ deg}^{-1}. \qquad \ldots\ldots(5)$$

If points below 2·5°K are included, (5) becomes

$$C = 0{\cdot}000346T + 464{\cdot}5(T/159{\cdot}0)^3 \text{ cal mol}^{-1}\text{ deg}^{-1} \qquad \ldots\ldots(5\,a)$$

indicating that the adsorbed helium was having some effect, though small.

Values of Θ, calculated after subtraction of the electronic contribution taken from (5), are shown in figure 5.

(iii) *Potassium*

Seven series of measurements were made on the potassium sample. This large number was necessary because a thermal e.m.f., with an associated long equilibrium time, developed in the thermometer leads, invalidating a large number of points. New leads fitted after series 5 eliminated the effect and a monotonically increasing atomic heat curve was obtained. Smoothed values are given in table 1.

A plot of C/T against T^2 is shown in figure 4. The behaviour is clearly non-linear at least as low as 2°K, at which temperature the points are known to be unreliable owing to helium adsorption. An exact analysis into cubic and linear terms is therefore not possible. However, putting in an approximate correction for the adsorbed helium and using points below 2°K, a linear term of about $0{\cdot}00047T$ cal mol^{-1} deg^{-1} is obtained, together with a lattice term equivalent to a Debye temperature of 89°K. The values of Θ shown in figure 5 were calculated after subtraction of the above linear term. With no correction the value of the latter is $0{\cdot}00058T$.

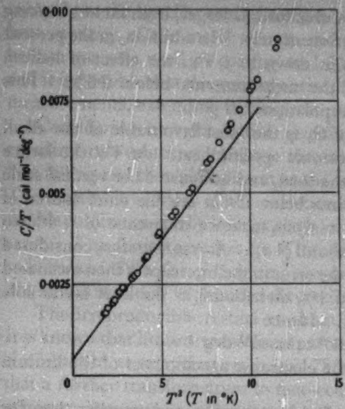


Figure 4. Potassium. The full line corresponds to $C=0{\cdot}00047T+464{\cdot}5(T/89)^3$ cal mol^{-1} deg^{-1}.

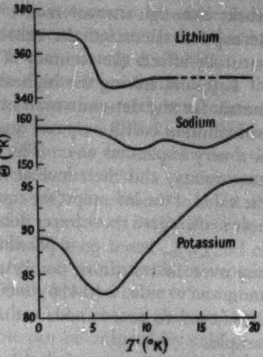


Figure 5. Debye temperatures.

(iv) *Note on Adsorbed Helium*

When the heat capacity of the calorimeter, containing only helium thermal exchange gas, was originally measured, no effect due to adsorption of the gas was noticeable. However, at a later date, check measurements showed a much increased heat capacity below 2·5°K. The excess over the original heat capacity, integrated with respect to temperature, was found to correspond to a total quantity of heat of 2×10^{-4} joule, and it disappeared on pumping the gas from the inside of the calorimeter. It was therefore thought to have been caused by the desorption of helium from the walls.

For comparison a rough calculation was made, based on figures for the heat of adsorption of helium on glass (Keesom 1942), and it was found that an adsorbing area of about 240 cm^2 would be required to account for the observed quantity of heat. The apparent internal surface area of the calorimeter was only 80 cm^2, and the effect probably became observable because of an increase in the effective area caused by oxidation.

The excess heat capacity, obtained by subtracting from the experimental points for lithium values calculated from (4), is very similar to the excess for the calorimeter, and can therefore be attributed to the same cause.

§ 5. DISCUSSION

The primary object of the present work was to obtain values for the electronic contribution to the low temperature specific heats of the alkali metals. This has been done by extrapolating to zero temperature, plots of C/T against T^2. Hence the accuracy with which the electronic term can be obtained depends on a number of factors other than the experimental accuracy of individual specific heat points. In particular, the range over which the C/T against T^2 plot can be expected to be linear depends on the Debye temperature of the metal, and is probably restricted to temperatures less than about $\Theta/50$ (Blackman 1955 a). Also the cubic term in the atomic heat is itself proportional to Θ^{-3}. Hence, if Θ is small, not only is the linear range more restricted, but the cubic term is larger, both factors making it more difficult to estimate the linear term accurately. In addition, in the present work, although measurements were taken down to 1·4°K, the effect of helium desorption decreased the reliability of the measurements below 2·5°K. This naturally affects the accuracy of the extrapolation.

Lithium, having the highest value of Θ, is the most favourable of the alkali metals for the determination of the electronic specific heat. At 4°K the lattice contribution is still only one quarter of the total, and in figure 1 the vertical scale is a very expanded one. The high points below 2·4°K are the chief source of uncertainty, and their inclusion in the analysis makes a difference of $+4\%$ to the value of the intercept (see equations (4) and (4 *a*)). It was therefore considered best to disregard these lower points, and the error in the intercept is then estimated to be $\pm 2\%$, based on a possible bias of 1% distributed in the least favourable way over the remaining points up to 4°K. Hence

$$(0{\cdot}418 \pm 0{\cdot}008)T \times 10^{-3} \text{ cal mol}^{-1} \text{ deg}^{-1}$$

is believed to be a reliable estimate of the electronic specific heat of the lithium sample.

In the sodium results, although the Debye temperature is smaller than for lithium, the linear range appears to extend to as high or higher temperatures. However at 1·7°K the lattice specific heat is already as great as that of the electrons. Helium desorption makes a difference of $5\frac{1}{2}\%$ in the intercept (equations (5) and (5 *a*)), but to ignore all points below 2·4°K and proceed as for lithium, results in an even larger and obviously erroneous estimate of the error in the intercept. Therefore, using the information obtained from the empty calorimeter and from the lithium results (see § 4 (iv)), the lowest temperature points have been corrected approximately for desorbed helium, and from the remaining uncertainty in these points the probable error in the linear term of (5) is estimated to be $\pm 3\%$. This gives $(0{\cdot}328 \pm 0{\cdot}01)T \times 10^{-3}$ cal mol^{-1} deg^{-1} for the electronic specific heat of the sodium sample.

In potassium, for which $\Theta/50$ is about 1·8°K, it is not surprising that a reliable linear portion of graph is unavailable. However it seems reasonable to suppose that the intercept obtained from the uncorrected points below 2°K ($0{\cdot}58 \times 10^{-3}$ cal mol^{-1} deg^{-2}) is too large, and that the corrected value of $0{\cdot}47 \times 10^{-3}$ cal mol^{-1} deg^{-2} is a better approximation to the true magnitude.

If the above values of the electronic specific heats are converted into effective masses, using equations (1) and (2) and assuming one electron per atom, the following results are obtained: for lithium $m^* = 2{\cdot}32m_0$, for sodium $m^* = 1{\cdot}22m_0$, and for potassium $m^* = 1{\cdot}1m_0$. All these values are greater than unity, in qualitative agreement with the soft x-ray work of Skinner (1940) on lithium and sodium. The reason may be simply the effect on the electrons of the periodic structure of the lattice (Mott and Jones 1936), which is ignored in the Sommerfeld free-electron model. Alternatively there is a more recent suggestion based on electron–lattice interactions (Buckingham and Schafroth 1954), which also gives rise to an electronic specific heat greater than the Sommerfeld value. The only comparable experimental value for the low temperature electronic specific heat of an alkali metal is that of $0{\cdot}00043T$ cal mol^{-1} deg^{-1} ($m^* = 1{\cdot}6m_0$) for sodium (Parkinson and Quarrington 1955). This is significantly higher than the present value, but the discrepancy may be explained by probable differences in the crystal structure of the sample (see below). Rayne (1954) has reported an anomaly in the specific heat of sodium below 1°K. In his method the combined heat capacity of the metal sample and demagnetization salt was measured and helium gas was used in cooling to 1·3°K. Both these features are undesirable and it is possible that the anomaly arises from the experimental method. Further measurements, extending to temperatures below 1°K, are clearly desirable.

Considering now the lattice specific heats, it is seen from figure 5 that the experimental values of Θ show the usual general variation with temperature, first falling and then rising again as the temperature increases (Blackman 1955 b). The relative smallness of the variation of Θ for sodium is remarkable, and the double minimum is a little unexpected, but seems to lie just outside experimental error. The limiting values as T tends to zero, 369°K for lithium, 158°K for sodium and 89°K for potassium, are in reasonable agreement with the values 354°K, 143°K and 77°K obtained theoretically (Fuchs 1936).

The irreproducible results at about 15°K in lithium are difficult to explain. It is known that lithium transforms partially from body-centred cubic to hexagonal close-packed at temperatures below 78°K, and it has been shown (Barrett 1956) that a further transformation to face-centred cubic can be induced by cold work at lower temperatures. However there is no evidence of a spontaneous transformation near 15°K which might account for the present observations. An alternative explanation involving an electronic transition seems unlikely in an alkali metal. The value of 0·095 cal mol^{-1} deg^{-1} for the atomic heat at 20°K is in exact agreement with the observations of Simon and Swain (1935), and their value of 0·045 cal mol^{-1} deg^{-1} at 15°K lies in the middle of the present results.

In sodium, not only is the electronic specific heat smaller than that observed by Parkinson and Quarrington (1955), but also the atomic heat over the whole range from 1·4° to 20°K is between 5% and 10% smaller than theirs. On the other hand the agreement with the values of Simon and Zeidler (1926) at 16·95° and 20°K is very good. The differences are probably in the method of preparing the sodium, the present sample and that of Simon and Zeidler being melted and cast, whereas that of Parkinson and Quarrington was cut from a solid lump of metal. That discrepancies exist between measurements on different samples is not surprising in view of the recent work of Barrett (1956). This shows that a partial crystallographic transformation from body-centred cubic to hexagonal close-packed takes place below 36°K, the fraction of the sample transformed depending on the purity and state of strain of the metal and the temperature to which it has been cooled. It therefore seems probable that greater consistency between the results of different experimenters will only be obtained when crystal structure studies can accompany the specific heat measurements, or alternatively, if some method of inhibiting the transformation can be evolved.

The values obtained for the atomic heat of potassium also can be compared with measurements of Simon and Zeidler (1926) above 15°K, the present values being 3% to 10% higher than theirs. Potassium does not show any crystallographic transformations (Barrett 1956).

§ 6. CONCLUSION

The work reported here must in some ways be regarded as a preliminary investigation only. In measurements of the atomic heats of lithium and sodium more information, obtained preferably under the same conditions as the atomic heat, is required on the crystal structures of the samples used. Also it is desirable, particularly with sodium and potassium, for the measurements to be extended to temperatures below 1°K in an apparatus which does not require the use of helium gas for thermal exchange. Measurements on caesium and rubidium could then also be considered.

ACKNOWLEDGMENTS

The author is indebted to Dr. J. C. Chaston of Johnson, Matthey Research Laboratories for the lithium sample, and to Dr. P. C. L. Pfeil of the Atomic Energy Research Establishment, Harwell, for the sodium and potassium samples. She also wishes to thank Mr. D. C. Handley for his invaluable assistance in making the measurements, Mrs. D. G. Maisey for her help in the use of the electronic computer for the least squares analyses, and Dr. D. H. Parkinson for his advice, encouragement and continued interest in this work. This paper is published by permission of the Controller, H.M. Stationery Office.

REFERENCES

BARRETT, C. S., 1956, *Acta Cryst., Camb.*, **9**, 671.
BLACKMAN, M., 1955 a, *Handb. Phys.*, **7**, Teil 1, p. 365 ; 1955 b, *Ibid.*, p. 361–372.
BUCKINGHAM, M. J., and SCHAFROTH, M. R., 1954, *Proc. Phys. Soc.* A, **67**, 828.
CLUSIUS, K., and GOLDMAN, J., 1936, *Z. Phys. Chem.*, **31**, 256.
FUCHS, K., 1936, *Proc. Roy. Soc.* A, **157**, 444.
KEESOM, W. H., 1942, *Helium* (Amsterdam : Elsevier), p. 127.
KEESOM, W. H., and CLARK, C. W., 1935, *Physica*, **2**, 513.
MOTT, N. F., and JONES, H., 1936, *Properties of Metals and Alloys* (Oxford : Clarendon Press), p. 81.
PARKINSON, D. H., 1954, *J. Sci. Instrum.*, **31**, 178.
PARKINSON, D. H., and QUARRINGTON, J. E., 1955, *Proc. Phys. Soc.* A, **68**, 762.
PARKINSON, D. H., and ROBERTS, L. M., 1957, *Proc. Phys. Soc.* B, **70**, 471.
RAYNE, J., 1954, *Phys. Rev.*, **95**, 1428.
SIMON, F., and SWAIN, R. C., 1935, 2, *Phys. Chem.* B, **28**, 189.
SIMON, F., and ZEIDLER, W., 1926, *Z. Phys. Chem.*, **123**, 383.
SKINNER, H. W. B., 1940, *Phil. Trans. Roy. Soc.* A, **239**, 12.
WILSON, A. H., 1936, *The Theory of Metals* (Cambridge : University Press), p. 185.

The Influence of a Strong Magnetic Field on the Dielectric Constant of a Diamagnetic Fluid

By A. D. BUCKINGHAM
Physical Chemistry Laboratory, University of Oxford

MS. received 14*th February* 1957, *and in revised form* 21*st March* 1957

Abstract. The effect of a strong magnetic field on the static dielectric constant of a diamagnetic fluid is investigated using statistical mechanical methods; a formula is derived relating the dielectric constant ϵ to the square of the field strength. The theory indicates that the change in ϵ is very small in non-polar substances (although it is detectable in magnetic double refraction experiments), but should be measurable in polar liquids; it is known to be substantial in 'liquid crystals', and it is suggested that it might also be large in solutions of large molecules. In a polar substance the chief contributor to the change in ϵ is related to the anisotropy in the diamagnetic susceptibility tensor and is proportional to the square of the molecular electric dipole moment. Relationships between the effect and electric and magnetic double refraction are established. Experimental data is scanty, but it is indicated how observations of the effect could lead to information about molecular properties.

---

§ 1. Introduction

The influence of a strong magnetic field on the refractive index of a fluid can be measured by observing its magnetic double refraction, that is, its Cotton–Mouton constant (Beams 1932). The effect is due to the favouring of certain molecular orientations by the magnetic field, leading in the case of anisotropically polarizable molecules to the macroscopic double refraction. The problem of the effect of a strong magnetic field on the static dielectric constant of a substance has not received much attention either by experimentalists or by theoreticians. Van Vleck (1932) discussed some relevant data and deduced a formula applicable to paramagnetic gases. It is clear that the change in the dielectric constant is normally very small. However, in liquid crystals large changes have been observed (Kast 1924, Bauer 1926, Marinin and Tsvetkov 1939), and appreciable effects have been found in some solids (McMahon 1956).

In a strong magnetic field H, molecules with anisotropic diamagnetic susceptibilities will favour certain orientations. Thus nitrobenzene will want to have the plane of its aromatic ring parallel to H, although there will not be a tendency for the electric dipoles either to be parallel or anti-parallel to the field; the situation therefore approaches that of a dipole in two dimensions. When a weak electric field is applied in the same direction as H, the parallel dipolar configuration will be favoured, and the electric polarization (and hence the dielectric constant) will depend on H.

Just as the Cotton–Mouton effect is analogous to the Kerr effect, so the dependence of the static dielectric constant ϵ on H is akin to dielectric saturation. In a strong uniform electric field E, ϵ can be written as a power series in E^2

$$\epsilon=\epsilon_0+bE^2+cE^4+\ldots. \qquad \ldots\ldots(1)$$

b and c are characteristic of the specimen, and it has been shown (Van Vleck 1932, Buckingham 1956) that

$$b=q_0+\frac{q_1}{kT}+\frac{q_2}{k^2T^2}+\frac{q_3}{k^3T^3}.$$

q_0 is related to the molecular 'hyperpolarizability', and q_1 is dependent upon the anisotropy in the polarizability tensor. The term in q_2 is also dependent on the anisotropy but in addition is proportional to the square of the molecular dipole moment μ. q_3 represents the dipole saturation term, and is proportional to μ^4.

In a diamagnetic material, a strong magnetic field will act similarly to E in favouring certain configurations. If we write

$$\epsilon=\epsilon_0+BH^2+CH^4+\ldots \qquad \ldots\ldots(2)$$

then B can be expanded like b, only in B there is no term in T^{-3}. The remainder of this paper is devoted to a discussion of B.

§ 2. THE GENERAL THEORY

For convenience, we consider a macroscopic spherical specimen of volume V containing N similar diamagnetic molecules. We suppose that this sphere is in a strong uniform field H; strictly H is the field *in vacuo*, but since the permeability of a diamagnetic material is nearly unity, the field in the medium is equal to H. The static dielectric constant may be defined by the equation

$$\frac{\epsilon-1}{\epsilon+2}E_0=\frac{4\pi}{3V}\overline{\mathbf{M}\cdot\mathbf{e}} \qquad \ldots\ldots(3)$$

where E_0 is the strength of a uniform electric field *in vacuo* (so that the field strength in the medium is $E=3E_0/(\epsilon+2)$), and $\overline{\mathbf{M}\cdot\mathbf{e}}$ is the mean value of the moment $\mathbf{M}$ of the specimen in the direction of the electric field, which we designate by the unit vector $\mathbf{e}$.

From (2) we obtain

$$\frac{\epsilon-1}{\epsilon+2}E_0=\frac{\epsilon_0-1}{\epsilon_0+2}E_0+\frac{3B}{(\epsilon_0+2)^2}H^2E_0+\ldots \qquad \ldots\ldots(4)$$

so that

$$B=\frac{2\pi(\epsilon_0+2)^2}{9}\left[\frac{d^3}{dE_0\,dH^2}\left(\frac{\overline{\mathbf{M}\cdot\mathbf{e}}}{V}\right)\right]_{E_0=0,H=0}. \qquad \ldots\ldots(5)$$

For liquids, V may depend upon E and H because of electrostriction and magnetostriction. V can be expanded as a power series

$$V=V_0+V\ \ H^2+V_EE^2+V_{EH}E^2H^2+\ldots, \qquad \ldots\ldots(6)$$

where V_H and V_E are in principle measurable quantities. Using thermodynamical reasoning similar to that of Scaife (1956) it can be shown that

$$V_H=\tfrac{1}{2}\left(\frac{\partial^2V}{\partial H^2}\right)_{E=0\,H=0},$$

$$=-\frac{1}{8\pi}\frac{\partial}{\partial P}\{(\mu^*-1)V\}=-\frac{V}{8\pi}\left\{\frac{\partial\mu^*}{\partial P}-(\mu^*-1)\beta\right\} \qquad \ldots\ldots(7)$$

where P is the pressure, μ^* the magnetic permeability and $\beta = -V^{-1}\partial V/\partial P$ the isothermal compressibility of the substance. From (5) and (6) we obtain

$$B = B_{\mathrm{D}} + B_{\mathrm{V}} = \frac{2\pi(\epsilon_0+2)^2}{9V}\left[\frac{d^3}{dE_0\,dH^2}\overline{(\mathbf{M}\,.\,\mathbf{e})} - \frac{2V_H}{V}\frac{d}{dE_0}\overline{(\mathbf{M}\,.\,\mathbf{e})}\right]_{E_0=0,\,H=0}. \qquad \text{......(8)}$$

B_{V}, which is the magnetostriction term, is therefore given by

$$B_{\mathrm{V}} = -(\epsilon_0-1)(\epsilon_0+2)\,V_H/3V. \qquad \text{......(9)}$$

If B is measured at constant volume, then $B_{\mathrm{V}}=0$ and $B=B_{\mathrm{D}}$.

If we assume that the configuration of the specimen can be represented by a continuous variable τ, classical statistical mechanics leads to the formula

$$\overline{\mathbf{M}\,.\,\mathbf{e}} = \frac{\int(\mathbf{M}\,.\,\mathbf{e})\exp\{-U(\tau,\mathbf{E}_0,\mathbf{H})/kT\}\,d\tau}{\int\exp\{-U(\tau,\mathbf{E}_0,\mathbf{H})/kT\}\,d\tau} \qquad \text{......(10)}$$

where $U(\tau, \mathbf{E}_0, \mathbf{H})$ is the potential energy of the sphere when its configuration is τ and the external fields $\mathbf{E}_0$ and $\mathbf{H}$; the integrations with respect to τ are over all permissible configurations. From (10) we obtain the result

$$\begin{aligned}\left[\frac{d^3}{dE_0\,dH^2}\overline{\overline{(\mathbf{M}\,.\,\mathbf{e})}}\right]_{E_0=0,\,H=0} = {} & \left\langle\frac{d^3(\mathbf{M}\,.\,\mathbf{e})}{dE_0\,dH^2}\right\rangle - \frac{1}{kT}\Biggl\langle\frac{d^2(\mathbf{M}\,.\,\mathbf{e})}{dH^2}\frac{dU}{dE_0} \\ & + \frac{d(\mathbf{M}\,.\,\mathbf{e})}{dE_0}\left\{\frac{d^2U}{dH^2} - \left\langle\frac{d^2U}{dH^2}\right\rangle\right\} + (\mathbf{M}\,.\,\mathbf{e})\left\langle\frac{d^3U}{dE_0\,dH^2}\right\rangle \\ & + \frac{1}{k^2T^2}\left\langle(\mathbf{M}\,.\,\mathbf{e})\frac{dU}{dE_0}\left\{\frac{d^2U}{dH^2} - \left\langle\frac{d^2U}{dH^2}\right\rangle\right\}\right\rangle\Biggr\rangle,\end{aligned} \qquad \text{......(11)}$$

since $\langle dU/dE_0\rangle = -\langle\mathbf{M}\,.\,\mathbf{e}\rangle = 0$, and $(dU/dH)_{H=0} = 0$ for diamagnetic materials. The angular brackets imply an average value in the absence of external fields, that is

$$\langle X\rangle = \frac{\int X(\tau,0,0)\exp\{-U(\tau,0,0)/kT\}\,d\tau}{\int\exp\{-U(\tau,0,0)/kT\}\,d\tau}. \qquad \text{......(12)}$$

$U(\tau, \mathbf{E}_0, \mathbf{H})$ will now be expanded as a power series in E_0 and H; using a tensor suffix notation

$$\begin{aligned}U = U_0 - M_{0_\alpha}E_{0_\alpha} - \tfrac{1}{2}A_{\alpha\beta}E_{0_\alpha}E_{0_\beta} - \tfrac{1}{2}X_{\alpha\beta}H_\alpha H_\beta \\ - \tfrac{1}{2}Y_{\alpha:\beta\gamma}E_{0_\alpha}H_\beta H_\gamma - \tfrac{1}{4}Z_{\alpha\beta:\gamma\delta}E_{0_\alpha}E_{0_\beta}H_\gamma H_\delta - \dots\ .\end{aligned} \qquad \text{......(13)}$$

From the general result

$$M_\alpha = -dU/dE_{0_\alpha} \qquad \text{......(14)}$$

we obtain

$$M_\alpha = M_{0_\alpha} + A_{\alpha\beta}E_{0_\beta} + \tfrac{1}{2}Y_{\alpha:\beta\gamma}H_\beta H_\gamma + \tfrac{1}{2}Z_{\alpha\beta:\gamma\delta}E_{0_\beta}H_\gamma H_\delta + \dots \qquad \text{......(15)}$$

so that M_{0_α} may be identified as the total electric dipole moment in the absence of external fields, $A_{\alpha\beta}$ with the electric polarizability tensor of the macroscopic sphere, $X_{\alpha\beta}$ with its diamagnetic susceptibility tensor and $Y_{\alpha:\beta\gamma}$ and $Z_{\alpha\beta:\gamma\delta}$ describe the effects of a magnetic field on the electric dipole moment. The expansions (13) and (15) are analogous to those used elsewhere (Buckingham and Pople 1955, 1956). It is obvious that $A_{\alpha\beta}$ and $X_{\alpha\beta}$ are symmetric in α and β, $Y_{\alpha:\beta\gamma}$ in β and γ, while $Z_{\alpha\beta:\gamma\delta}$ is symmetric in the separate pairs α, β and γ, δ.

From (8), (11), (13) and (14),

$$B_{\mathrm{D}}=\frac{2\pi(\epsilon_0+2)^2}{9V}\left\langle\left\{Z_{\alpha\beta:\gamma\delta}+\frac{2}{kT}M_{0_\alpha}Y_{\beta:\gamma\delta}\right.\right.$$

$$\left.+\frac{1}{kT}\left(A_{\alpha\beta}+\frac{1}{kT}M_{0_\alpha}M_{0_\beta}\right)X_{\gamma\delta}\right\}e_\alpha e_\beta h_\gamma h_\delta$$

$$\left.-\frac{1}{kT}\left(A_{\alpha\beta}+\frac{1}{kT}M_{0_\alpha}M_{0_\beta}\right)e_\alpha e_\beta\langle X_{\gamma\delta}h_\gamma h_\delta\rangle\right\rangle. \qquad \ldots\ldots(16)$$

The averaging over all molecular configurations relative to **e** and **h** can now be performed. We shall represent the angle between $\mathbf{E}_0$ and $\mathbf{H}$ by Ω, so that $e_\alpha h_\alpha=\cos\Omega$. By keeping Ω fixed and averaging over all directions of **e** and **h**, we obtain

$$\langle e_\alpha e_\beta\rangle=\langle h_\alpha h_\beta\rangle=\tfrac{1}{3}\delta_{\alpha\beta}$$

$$\langle e_\alpha e_\beta h_\gamma h_\delta\rangle=\frac{1}{9}\delta_{\alpha\beta}\delta_{\gamma\delta}+\frac{1}{90}(3\cos^2\Omega-1)\{-2\delta_{\alpha\beta}\delta_{\gamma\delta}+3\delta_{\alpha\gamma}\delta_{\beta\delta}+3\delta_{\alpha\delta}\delta_{\beta\gamma}\}. \qquad \ldots\ldots(17)$$

Thus (16) becomes

$$B_{\mathrm{D}}=\frac{2\pi(\epsilon_0+2)^2}{405\,V}\left[3(3\cos^2\Omega-1)\left\langle Z_{\alpha\beta:\alpha\beta}-\tfrac{1}{3}Z_{\alpha\alpha:\beta\beta}+\frac{2}{kT}\left(M_{0_\alpha}Y_{\beta:\alpha\beta}-\tfrac{1}{3}M_{0_\alpha}Y_{\alpha:\beta\beta}\right)\right.\right.$$

$$\left.+\frac{1}{kT}\left\{\left(A_{\alpha\beta}+\frac{1}{kT}M_{0_\alpha}M_{0_\beta}\right)X_{\alpha\beta}-\tfrac{1}{3}\left(A_{\alpha\alpha}+\frac{1}{kT}M_{0_\alpha}M_{0_\alpha}\right)X_{\beta\beta}\right\}\right\rangle$$

$$\left.+5\left\langle Z_{\alpha\alpha:\beta\beta}+\frac{2}{kT}M_{0_\alpha}Y_{\alpha:\beta\beta}+\frac{1}{kT}\left(A_{\alpha\alpha}+\frac{1}{kT}M_{0_\alpha}M_{0_\alpha}\right)(X_{\beta\beta}-\langle X_{\beta\beta}\rangle)\right\rangle\right], \qquad \ldots\ldots(18)$$

and this is the general expression for B_{D}. To proceed further, we shall expand M_{0_α}, $A_{\alpha\beta}$, $X_{\alpha\beta}$, $Y_{\alpha:\beta\gamma}$ and $Z_{\alpha\beta:\gamma\delta}$ as sums over the individual molecules of the specimen. Using the expansions of Buckingham and Pople (1955, 1956) (except for factors of $\frac{1}{2}$ in the tensors ξ and η)

$$M_{0_\alpha}=\sum_i m_\alpha^{(i)}=\sum_i\{\mu_\alpha^{(i)}+\alpha_{\alpha\beta}^{(i)}F_\beta^{(i)}+\ldots\} \qquad \ldots\ldots(19)$$

$$A_{\alpha\beta}=\sum_i\frac{\partial m_\alpha^{(i)}}{\partial E_{0_\beta}}=\sum_i\{\alpha_{\alpha\beta}^{(i)}+\beta_{\alpha\beta\gamma}^{(i)}F_\gamma^{(i)}+\ldots\} \qquad \ldots\ldots(20)$$

$$X_{\alpha\beta}=\sum_i\{\chi_{\alpha\beta}^{(i)}+\xi_{\gamma:\alpha\beta}^{(i)}F_\gamma^{(i)}+\ldots\} \qquad \ldots\ldots(21)$$

$$Y_{\alpha:\beta\gamma}=\sum_i\{\xi_{\alpha:\beta\gamma}^{(i)}+\eta_{\delta\alpha:\beta\gamma}^{(i)}F_\delta^{(i)}+\ldots\} \qquad \ldots\ldots(22)$$

$$Z_{\alpha\beta:\gamma\delta}=\sum_i\{\eta_{\alpha\beta:\gamma\delta}^{(i)}+\ldots\}. \qquad \ldots\ldots(23)$$

The tensors μ, α, β, χ, ξ, η, … are characteristic properties of isolated molecules; they describe the energy u of a diamagnetic molecule in uniform electric and magnetic fields F_α and H_α

$$u=u^{(0)}-\mu_\alpha F_\alpha-\tfrac{1}{2}\alpha_{\alpha\beta}F_\alpha F_\beta-\tfrac{1}{6}\beta_{\alpha\beta\gamma}F_\alpha F_\beta F_\gamma-\tfrac{1}{2}\chi_{\alpha\beta}H_\alpha H_\beta-\tfrac{1}{2}\xi_{\alpha:\beta\gamma}F_\alpha H_\beta H_\gamma$$

$$-\tfrac{1}{4}\eta_{\alpha\beta:\gamma\delta}F_\alpha F_\beta H_\gamma H_\delta-\ldots, \qquad \ldots\ldots(24)$$

so that μ_α is the permanent dipole moment of an isolated molecule, $\alpha_{\alpha\beta}$ its polarizability tensor, $\chi_{\alpha\beta}$ its diamagnetic susceptibility tensor, and $\beta_{\alpha\beta\gamma}$, $\xi_{\alpha:\beta\gamma}$, $\eta_{\alpha\beta:\gamma\delta}$

describe the non-linear terms relating m to F and H. In (19)–(23), $F_\alpha^{(i)}$ is the electric field acting on molecule i due to the charge distribution of its neighbours ($F_\alpha^{(i)}$ is assumed constant over the volume occupied by the ith molecule).

On substituting (19)–(23) into (18), and after retaining leading terms only, we obtain

$$B_{\mathrm{D}}=\frac{2\pi N(\epsilon_0+2)^2}{405\,V}\Big[3(3\cos^2\Omega-1)\Big\{\eta_{\alpha\beta:\alpha\beta}{}^{(1)}-\tfrac{1}{3}\eta_{\alpha\alpha\ \beta\beta}{}^{(1}$$
$$+\frac{1}{kT}\Big\langle 2\xi_{\beta:\alpha\beta}{}^{(1)}M_{0_\alpha}-\tfrac{2}{3}\xi_{\alpha:\beta\beta}{}^{(1)}M_{0_\alpha}+\chi_{\alpha\beta}{}^{(1)}A_{\alpha\beta}-\tfrac{1}{3}\chi_{\alpha\alpha}{}^{(1)}A_{\beta\beta}$$
$$+\frac{1}{kT}(\chi_{\alpha\beta}{}^{(1)}M_{0_\alpha}M_{0_\beta}-\tfrac{1}{3}\chi_{\alpha\alpha}{}^{(1)}M_{0_\beta}M_{0_\beta})\Big\rangle\Big\}+5\eta_{\alpha\alpha:\beta\beta}{}^{(1)}+\frac{10}{kT}\langle\xi_{\alpha:\beta\beta}{}^{(1)}M_{0_\alpha}\rangle\Big]. \quad \ldots\ldots(25)$$

There is a close similarity between several of the terms in (25) and those occurring in the corresponding expression for the molecular Kerr constant (Buckingham 1955, Buckingham and Raab 1957). When molecule 1 is fixed, the average value of M_{0_α} is $\{9\epsilon_0/(2\epsilon_0+1)(\epsilon_0+2)\}\overline{m_\alpha{}^{(1)}}$ (Kirkwood 1939) where $\overline{m_\alpha{}^{(1)}}$ is the mean dipole moment of molecule 1 and its neighbours in a macroscopic sphere small relative to V. Also, when the principal axes of $\alpha_{\alpha\beta}$ and $\chi_{\alpha\beta}$ coincide, to a good approximation

$$\langle\chi_{\alpha\beta}{}^{(1)}A_{\alpha\beta}-\tfrac{1}{3}\chi_{\alpha\alpha}{}^{(1)}A_{\beta\beta}\rangle=\chi_{\alpha\beta}\alpha_{\alpha\beta}-3\chi\alpha \quad \ldots\ldots(26)$$

where $\alpha=\frac{1}{3}\alpha_{\alpha\alpha}$ and $\chi=\frac{1}{3}\chi_{\alpha\alpha}$ are the mean polarizability and diamagnetic susceptibility of an isolated molecule. It is convenient to define

$$\kappa_\alpha{}^2=\frac{\alpha_{\alpha\beta}\alpha_{\alpha\beta}-3\alpha^2}{6\alpha^2},\quad \kappa_\chi{}^2=\frac{\chi_{\alpha\beta}\chi_{\alpha\beta}-3\chi^2}{6\chi^2}. \quad \ldots\ldots(27)$$

For axially symmetric molecules with polarizabilities $\alpha_{\parallel}$ and $\alpha_{\perp}$ along and at right angles to the axis $\kappa_\alpha=(\alpha_{\parallel}-\alpha_{\perp})/3\alpha$, $\kappa_\chi=(\chi_{\parallel}-\chi_{\perp})/3\chi$. κ_α and κ_χ are the *anisotropies* in $\alpha_{\alpha\beta}$ and $\chi_{\alpha\beta}$ respectively. In (26) terms in $\Sigma_{i\neq 1}\langle\cos^2\theta^{(1i)}-\frac{1}{3}\rangle$, where $\theta^{(1i)}$ is the angle between the axes of molecules 1 and i, have been omitted, but in the abovementioned papers on the Kerr effect they were shown usually to be negligible.

For axially symmetric molecules

$$\langle\chi_{\alpha\beta}{}^{(1)}M_{0_\alpha}M_{0_\beta}-\tfrac{1}{3}\chi_{\alpha\alpha}{}^{(1)}M_{0_\beta}M_{0_\beta}\rangle=(\chi_{\parallel}-\chi_{\perp})\langle M_1{}^2-\tfrac{1}{3}M^2\rangle=3\kappa_\chi\chi f \quad \ldots\ldots(28)$$

where $f=\langle M_1{}^2-\frac{1}{3}M^2\rangle$ has been discussed in detail elsewhere (Buckingham and Raab 1957); $M_1{}^2=\mu^{-2}\mu_\alpha{}^{(1)}M_{0_\alpha}\mu_\beta{}^{(1)}M_{0_\beta}$ is the square of the component of $\mathbf{M}_0$ in the direction of $\boldsymbol{\mu}^{(1)}$. Thus (25) reduces to

$$B_{\mathrm{D}}=\frac{2\pi N(\epsilon_0+2)^2}{405\,V}\Big[3(3\cos^2\Omega-1)\Big\{\eta_{\alpha\beta:\alpha\beta}-\tfrac{1}{3}\eta_{\alpha\alpha:\beta\beta}$$
$$+\frac{2}{kT}\frac{9\epsilon_0}{(2\epsilon_0+1)(\epsilon_0+2)}(\xi_{\beta:\alpha\beta}-\tfrac{1}{3}\xi_{\alpha:\beta\beta})\overline{m_\alpha}+\frac{3}{kT}\kappa_\chi\chi\Big(2\kappa_\alpha\alpha+\frac{f}{kT}\Big)\Big\}$$
$$+5\eta_{\alpha\alpha:\beta\beta}+\frac{10}{kT}\frac{9\epsilon_0}{(2\epsilon_0+1)(\epsilon_0+2)}\xi_{\alpha:\beta\beta}\overline{m_\alpha}\Big]. \quad \ldots\ldots(29)$$

Two special cases are of interest, viz. when $\mathbf{E}_0$ and $\mathbf{H}$ are parallel so that $\cos^2\Omega=1$ and when they are perpendicular when $\cos\Omega=0$. From (29) we calculate that

$$B^{\parallel}-B^{\perp}=B_{\mathrm{D}}{}^{\parallel}-B_{\mathrm{D}}{}^{\perp}$$
$$=\frac{2\pi N(\epsilon_0+2)^2}{45\,V}\Big[\eta_{\alpha\beta:\alpha\beta}-\tfrac{1}{3}\eta_{\alpha\alpha:\beta\beta}+\frac{2}{kT}\frac{9\epsilon_0}{(2\epsilon_0+1)(\epsilon_0+2)}$$
$$+(\xi_{\beta:\alpha\beta}-\tfrac{1}{3}\xi_{\alpha:\beta\beta})\overline{m_\alpha}+\frac{3}{kT}\kappa_\chi\chi\Big(2\kappa_\alpha\alpha+\frac{f}{kT}\Big)\Big] \quad \ldots\ldots(30)$$

and this quantity has the advantage of being independent of the magnetostriction term B_V.

§ 3. SPECIAL CASES

3.1. *Non-dipolar Molecules*

For molecules with centres of inversion, $\mu_\alpha = \beta_{\alpha\beta\gamma} = \xi_{\alpha:\beta\gamma} = 0$, and f is negligible (although it may not be zero, for molecular electric quadrupole moments can induce dipoles). Thus (29) in this case becomes

$$B_D = \frac{2\pi N(\epsilon_0+2)^2}{405\,V}\left[3(3\cos^2\Omega-1)\left\{\eta_{\alpha\beta:\alpha\beta} - \tfrac{1}{3}\eta_{\alpha\alpha:\beta\beta} + \frac{6}{kT}\kappa_\alpha\kappa_\chi{}^\alpha\chi\right\} + 5\eta_{\alpha\alpha:\beta\beta}\right]. \qquad \ldots\ldots(31)$$

3.2. *Spherical Molecules*

For spherical molecules there are two independent η's,

$$\eta_\parallel = \eta_{11:11} \text{ and } \eta_\perp = \eta_{11:22} \quad (2\eta_{12:12} = \eta_\parallel - \eta_\perp),$$

so that (31) becomes

$$B_D = \frac{2\pi N(\epsilon_0+2)^2}{27\,V}[(3\cos^2\Omega-1)(\eta_\parallel-\eta_\perp)+\eta_\parallel+2\eta_\perp], \qquad \ldots\ldots(32)$$

and in this case B_D is proportional to the density, for interaction terms will be small. From (32)

$$B_D{}^\parallel = \frac{2\pi N(\epsilon_0+2)^2}{9\,V}\eta_\parallel$$

$$B_D{}^\perp = \frac{2\pi N(\epsilon_0+2)^2}{9\,V}\eta_\perp. \qquad \ldots\ldots(33)$$

3.3. *Polar Anisotropic Molecules*

If the molecules are strongly polar and in addition have appreciable diamagnetic anisotropies (e.g. nitrobenzene), then the dominant terms in (29) are those in κ_χ, so that for them

$$B_D = \frac{2\pi N(\epsilon_0+2)^2}{45\,kTV}\kappa_\chi\chi\left(2\kappa_\alpha\alpha + \frac{f}{kT}\right)(3\cos^2\Omega-1). \qquad \ldots\ldots(34)$$

For the Onsager (1936) model of a polar liquid it has been shown (Buckingham and Raab 1957) that

$$f = \frac{27\,kTV}{2\pi N}\frac{\epsilon_0(\epsilon_0-\epsilon_\infty)}{(\epsilon_0+2)^2(2\epsilon_0+\epsilon_\infty)} \qquad \ldots\ldots(35)$$

where ϵ_∞ is the high-frequency dielectric constant of the substance, that is,

$$\frac{\epsilon_\infty-1}{\epsilon_\infty+2} = \frac{4\pi N\alpha}{3V}. \qquad \ldots\ldots(36)$$

§ 4. DISCUSSION

There is very little evidence about the magnitude of the η tensor. For atomic hydrogen, an exact calculation (Buckingham and Pople 1957) (neglecting the effects of electron spin) gives

$$\eta_\parallel = -\frac{159}{8}\frac{a_0{}^6}{mc^2}, \quad \eta_\perp = -\frac{797}{24}\frac{a_0{}^6}{mc^2} \qquad \ldots\ldots(37)$$

when energies are in ergs, E_0 n e.s.u. and H in oersteds; a_0 is the Bohr radius

radius (0·5292 Å), m the electronic mass and c the velocity of light. Thus $\eta_{\parallel} = -53 \times 10^{-44}$, $\eta_{\perp} = -89 \times 10^{-44}$. The experimental value (Buckingham and Pople 1956) of $\eta_{\alpha\beta:\alpha\beta} - \frac{1}{3}\eta_{\alpha\alpha:\beta\beta}$ for 1 : 3 : 5 trimethylbenzene is $-3{\cdot}5 \times 10^{-40}$, and for a sphere $\eta_{\alpha\beta:\alpha\beta} - \frac{1}{3}\eta_{\alpha\alpha:\beta\beta} = 5\,(\eta_{\parallel} - \eta_{\perp})$. For carbon tetrachloride $\eta_{\parallel}$ and $\eta_{\perp}$ might be about thirty times as large as the corresponding atomic hydrogen constants (this is the ratio of their electric hyperpolarizabilities), so that, from (33), for CCl_4 at S.T.P., $B_{\mathrm{D}}{}'' = -1{\cdot}3 \times 10^{-18}$. Thus for a field of 2×10^4 oersteds $\Delta\epsilon = \epsilon - \epsilon_0 = -5 \times 10^{-10}$, which is far too small to be measurable with the normal techniques.

For liquid nitrobenzene Piekara (1936) has reported that $\Delta\epsilon$ is 6×10^{-4} for $H = 20\,400$ oersteds (Ω is not stipulated). B_{D} for this liquid can be estimated through (34) and (35). On using $\kappa_{\alpha}{}^2 = (\alpha_1{}^2 + \alpha_2{}^2 + \alpha_3{}^2 - 3\alpha^2)/6\alpha^2$, where α_1 is the polarizability in the direction of the permanent molecular dipole and α_2 and α_3 the other principal polarizabilities, the figures of Le Fèvre and Le Fèvre (1953) lead to $\kappa_{\alpha}\alpha = 3{\cdot}2 \times 10^{-24}\,\mathrm{cm}^3$. To determine $\kappa_{\chi}\chi$, the observed Cotton–Mouton constant (${}_{\mathrm{m}}C = 7 \times 10^{-16}$, Partington 1953) can be used with the equation (Buckingham and Pople 1956)

$$ {}_{\mathrm{m}}C = \frac{4\pi N}{45\,kT}\kappa_{\alpha}\kappa_{\chi}\alpha\chi \qquad \ldots\ldots(38)$$

which omits terms in η; N is Avogadro's number. Thus $\kappa_{\chi}\chi = 0{\cdot}54 \times 10^{-28}$. Hence, with $\epsilon_0 = 34{\cdot}9$, $\epsilon_{\infty} = 2{\cdot}63$, $B_{\mathrm{D}} = 2{\cdot}3 \times 10^{-14}\,(3\cos^2\Omega - 1)$, and for $H = 20\,400$ oersteds $\Delta\epsilon = 1{\cdot}0 \times 10^{-5}(3\cos^2\Omega - 1)$ at constant volume. For $\cos^2\Omega = 1$ this change in ϵ is only one thirtieth of Piekara's observed result, but this author admits that conductivity effects made his observations uncertain.

It is of interest to note that (30) is related to the expression for the molecular Kerr constant ${}_{\mathrm{m}}K$ of a polar substance (Buckingham 1955). On neglecting hyperpolarizability (normally a legitimate procedure for anisotropic polar molecules)

$$ {}_{\mathrm{m}}K = \frac{2\pi N\kappa_a a}{45\,kT}\left[2\kappa_{\alpha}\alpha + \frac{f}{kT}\right] \qquad \ldots\ldots(39)$$

where $3\kappa_a a$ is equal to $a_{\parallel} - a_{\perp}$, a being the polarizability for the frequency with which the Kerr constant is measured. From (30) and (39)

$$ {}_{\mathrm{m}}K = (B^{\parallel} - B^{\perp})\frac{\kappa_a a V_{\mathrm{m}}}{3\kappa_{\chi}\chi(\epsilon_0 + 2)^2} \qquad \ldots\ldots(40)$$

where V_{m} is the molar volume; (40) should be applicable at all densities. The above estimate of B_{D} for nitrobenzene at 25°C is confirmed by (40), for the observed value of ${}_{\mathrm{m}}K$ is 71×10^{-12} (Le Fèvre and Le Fèvre 1953).

The above calculations imply that B_{D} should be measurable with existing techniques in the case of polar liquids, and that data would be useful for the determination of diamagnetic susceptibility tensors $\chi_{\alpha\beta}$. For very large molecules (like polymers and viruses) with huge dipole moments $\Delta\epsilon$ should be measurable with ease and precision. By interpreting observations on dilute solutions of these 'macro-molecules' through (34), and therefore in terms of the solute's molecular anisotropy, it should be possible to learn something of the structure of the molecule itself. In liquid crystals there exist data (see § 1), and if these were combined with Kerr constant measurements, (40) would lead to a knowledge of the ratio of the electrical to the diamagnetic polarizability anisotropy and hence to some insight into the structure of the immense molecular clusters (perhaps containing 10^5 or 10^6 molecules) existing in these substances (see the Faraday Society 1933).

Since the magnetic field a few ångströms from a paramagnetic molecule or ion can be very large (of the order of 10^4 oersteds) it might be expected that there would be some saturation due to this field in the neighbourhood of the particle. However, the saturation arising from the electric field (which may be about 10^6 e.s.u. near an ion or dipole) would normally be much more important.

In conclusion we note that if the magnetic permeability μ^* is expanded similarly to (2),

$$\mu^* = \mu_0^* + \mathscr{B}E^2 + \mathscr{C}E^4 + \ldots$$

then $\mathscr{B} = \mathscr{B}_{\mathrm{D}} + \mathscr{B}_{\mathrm{V}}$, where $\mathscr{B}_{\mathrm{D}} = B_{\mathrm{D}}$ and $\mathscr{B}_{\mathrm{V}} = -(\mu_0^* - 1)\,V_E/V$. There is, therefore, a close connection between the effect of a magnetic field on ϵ and that of an electric field on μ^*.

Acknowledgment

The author is grateful to the Commissioners for the Exhibition of 1851 for the award of a Senior Studentship.

References

Bauer, E., 1926, *C.R. Acad. Sci., Paris*, **182**, 1541.
Beams, J. W., 1932, *Rev. Mod. Phys.*, **4**, 133.
Buckingham, A. D., 1955, *Proc. Phys. Soc.* A, **68**, 910; 1956, *J. Chem. Phys.*, **25**, 428.
Buckingham, A. D., and Pople, J. A., 1955, *Proc. Phys. Soc.* A, **68**, 905 ; 1956, *Proc. Phys. Soc.* B, **69**, 1133 ; 1957, *Proc. Camb. Phil. Soc.*, **53**, 262.
Buckingham, A. D., and Raab, R. E., 1957, *J. Chem. Soc.*, 2341.
Faraday Society, 1933, *Discussion on Liquid Crystals and Anisotropic Melts, Trans. Faraday Soc.*, **29**, 881.
Kast, W., 1924, *Ann. Phys., Lpz.*, **73**, 145.
Kirkwood, J. G., 1939, *J. Chem. Phys.*, **7**, 911.
Le Fèvre, C. G., and Le Fèvre, R. J. W., 1953, *J. Chem. Soc.*, 4041.
Marinin, V., and Tsvetkov, V., 1939, *Acta Physicochim., U.R.S.S.*, **11**, 837.
McMahon, W., 1956, *J. Amer. Chem. Soc.*, **78**, 3290.
Onsager, L., 1936, *J. Amer. Chem Soc.*, **58**, 1486.
Partington, J. R., 1953, *Advanced Treatise on Physical Chemistry*, Vol. IV (London : Longmans), p. 286.
Piekara, A., 1936, *C.R. Acad. Sci., Paris*, **202**, 206.
Scaife, B. K. P., 1956, *Proc. Phys. Soc.* B, **69**, 153.
Van Vleck, J. H., 1932, *The Theory of Electric and Magnetic Susceptibilities* (Oxford : University Press).

The High Temperature Dielectric Breakdown of Alkali Halides

By J. J. O'DWYER
Division of Electrotechnology, National Standards Laboratory, Sydney, N.S.W., Australia

MS. received 18*th December* 1956

Abstract. A model is proposed for the calculation of a critical field strength for an ionic crystal containing positively charged point defects which can act as electron trapping centres. It is assumed that conduction electrons gain energy from an applied field at a rate which is limited only by their collisions with lattice vibrations, and that this energy is transferred to the lattice mainly by multiphonon emission with recombination into empty trapping centres. A critical field is calculated and it is discussed under what circumstances it could be identified with the high-temperature dielectric breakdown field of alkali halides.

§ 1. Introduction

The general form of the variation of the dielectric breakdown strengths of alkali halides with temperature has been given by various authors (e.g. Austen and Whitehead 1940, Alger and von Hippel 1949, and Calderwood and Cooper 1953). It is the same for all the alkali halides measured. A maximum breakdown strength of about 10^6 volts cm^{-1} occurs in most cases at some transition temperature just above room temperature. Above and below this transition temperature are referred to as high and low temperature regions of breakdown.

Critical fields designed to explain breakdown in the low temperature region have been calculated by Fröhlich (1937, 1939, 1952), Callen (1949), O'Dwyer (1954) and Fröhlich and Paranjape (1956). All determine a critical field which is about the correct order of magnitude and has a similar temperature variation to that observed in the low temperature region.

It seems certain that crystal imperfections play a modifying role in the low temperature region, and that the results of the theories already mentioned should be merely corrected by a relatively small factor to allow for their presence. In the high temperature region they seem to play a decisive role—actually determining the breakdown mechanism. Fröhlich (1947) has examined a model in which electronic trapped states are closely spaced over a total energy width ΔV immediately below the conduction levels. Assuming that electron–electron collisions determine the distribution function of conduction electrons and trapped electrons, and that energy transfer from the electrons to the lattice is effected mainly by transitions between trapped states with the emission of phonons, Fröhlich finds a critical field strength

$$F^* = C\exp(\Delta V/2kT_0) \qquad \ldots\ldots(1)$$

where T_0 is the lattice temperature, and C is an undetermined constant with relatively slow temperature variation.

The application of (1) to the high temperature breakdown of alkali halides gives $\Delta V = 0{\cdot}12$ ev for NaCl and $\Delta V = 0{\cdot}073$ ev for KCl (cf. Calderwood and Cooper). These figures are too small (by a factor of about 10) for the ground states of electron traps due to the usual defects in alkali halides. Thus, unless it be argued that only the excited states of the trapping centres are to be considered, the Fröhlich amorphous dielectric theory does not apply well to the breakdown of alkali halides in their high temperature region.

It is the purpose of this paper to calculate a critical field strength for a model which seems to be more applicable to alkali halides in this region.

§ 2. THE MODEL

We consider an ionic crystal containing n_D similar positively charged point defects per unit volume which give rise to bound electron states below the conduction levels. It is assumed that n_D does not depend on the lattice temperature, although it undoubtedly must to some extent, and our assumption thus amounts to saying that the frozen-in defects greatly exceed those created thermally over the temperature range we consider.

For simplicity only the ground state of the electron in the trap is considered, and we assume that the total number of electrons in conduction levels and in traps is equal to the number of traps; furthermore these electrons may assume a temperature T (different from the lattice temperature T_0) under the influence of an applied field. The number of electrons in conduction levels (and also the number of vacant traps) is then given by (cf. Seitz 1940)

$$n_C = [n_D C (kT)^{3/2} \pi^{1/2}/2]^{1/2} \exp(-W/2kT) \qquad \ldots\ldots (2)$$

where

$$C = 4\pi (2m)^{3/2}/h^3$$

and W is the ionization energy of an electron in a trap (see footnote on page 764). The number of electrons of energy E above the bottom of the conduction band is then given by

$$n(E) = 2n_C (E/\pi k^3 T^3)^{1/2} \exp(-E/kT). \qquad \ldots\ldots (3)$$

Electron–electron collisions conserve the total electron momentum and have no effect on the rate of energy gain from an applied field. It will be assumed that this rate of energy gain is limited by the scattering of electrons by phonons, although in principle an additional term should be considered mainly due to scattering by ionized defects.

For the purpose of calculating a critical field strength the conduction electrons will be considered as losing energy to the lattice mainly by recombining with empty trapping centres by a multiphonon emission process. The theory of such radiationless recombination has been given by Huang and Rhys (1950) for alkali halides. A discussion on the appropriateness of this assumption is given below.

§ 3. CALCULATION OF THE CRITICAL FIELD STRENGTH

3.1. *Rate of Energy Gain of Conduction Electrons from an Applied Field*

The calculation of the rate at which a Maxwellian distribution of conduction electrons in a polar crystal gains energy from an applied field has been given by

Fröhlich (1952) when the rate of gain is hindered only by electron–phonon collisions. This result is

$$A = \frac{e^2 \tau F^2}{m} n_{\mathrm{C}} \qquad \ldots\ldots(4)$$

where e and m are the electronic charge and mass, F is the applied field and τ is a certain average relaxation time given by

$$\frac{1}{\tau} = \left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0}\right) \frac{e^2}{2\hbar^2} \left(\frac{\pi m \hbar \omega}{2}\right)^{1/2} x^{1/2} n(x_0) \{\exp(x_0 - x) + 1\} \qquad \ldots\ldots(5)$$

where ϵ_0 and ϵ_∞ are the static and high frequency dielectric constants, $\hbar\omega$ is the phonon energy (assumed to be all of the same circular frequency $\omega = (\epsilon_0/\epsilon_\infty)^{1/2} \omega_t$ where ω_t is the residual ray circular frequency),

$$x = \hbar\omega/kT, \quad x_0 = \hbar\omega/kT_0$$

and $n(x_0)$ is the Planck function.

3.2. *Rate of Energy Loss to the Lattice by the Conduction Electrons*

Huang and Rhys (1950) have shown that the probability of an electron making a transition between a trapped state μ and a conduction level $\mathbf{k}$ with multiphonon emission or absorption processes is given approximately by

$$p\begin{Bmatrix}\mathbf{k}, & \mu \\ \mu, & \mathbf{k}\end{Bmatrix} = \frac{8\pi^2 h\omega^2}{3a^3} \left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0}\right) \left\{n(x_0) + \tfrac{1}{2}\right\} \frac{|\langle \mathbf{k}|e\mathbf{x}|\mu\rangle|^2}{(W + \hbar k^2/2m)^2} R_{\pm p} \qquad \ldots\ldots(6)$$

where a is the interatomic distance, $\langle \mathbf{k}\,|e\mathbf{x}\,|\mu\rangle$ is the matrix element of the electric moment between states μ and $\mathbf{k}$, $\hbar^2k^2/2m$ is the energy of the conduction electron (assumed to be in a free electron state) and

$$R_p = \exp[\tfrac{1}{2}px_0 - S\{2n(x_0) + 1\}]\, I_p[2S\{n(x_0)[n(x_0) + 1]\}^{1/2}] \qquad \ldots\ldots(7)$$

where

$$p = (W + \hbar^2k^2/2m)/\hbar\omega. \qquad \ldots\ldots(8)$$

S is an interaction constant related to the equilibrium potential energy of the defect in the lattice, and $I_p(z)$ is the Bessel function of order p of imaginary argument. It should be noted that in (6) p is positive or negative according as we are considering recombination or ionization of the electron.

We can now calculate the rate of energy transfer from the electrons to the lattice as the excess of recombinations over ionizations multiplied by the energy involved and summed over all conduction levels. Thus since $P(\mu, \mathbf{k})$ and $P(\mathbf{k}, \mu)$ depend only on $|\mathbf{k}|$ we may consider a set of conduction levels of energy E, and the net rate of energy transfer to the lattice is given by (remembering that the density of empty traps is equal to the density of conduction electrons)

$$B(E) = (E + W)\{n(E)\, P(k, \mu)\, n_{\mathrm{C}} - n_{\mathrm{D}}\, P(\mu, k)\, \rho(E)\} \qquad \ldots\ldots(9)$$

where $\rho(E) = C\sqrt{E}$ is the density of conduction levels, and we have assumed that traps are almost all occupied and conduction levels almost all empty. From (6), (7) and (8)

$$\frac{P(\mu, k)}{P(k, \mu)} = \exp\{-(E + W)/kT_0\} \qquad \ldots\ldots(10)$$

since $I_p(z) = I_{-p}(z)$. Substituting for n_C in (9) we find $B(E)=0$ if $T=T_0$ so that the energy exchanges balance properly for the equilibrium state.†

Under the action of a strong field $T \neq T_0$ and the total energy transfer to the lattice is then

$$B = \int_0^\infty (E+W)\{n(E)P(k,\mu)n_C - n_D P(\mu,k)\rho(E)\}dE. \qquad \ldots\ldots(11)$$

Following Huang and Rhys we change the variable to p (cf. equation (8)), write $p_0 = W/\hbar\omega$, and neglect the dependence of $\langle \mathbf{k}|e\mathbf{x}|\mu\rangle$ on $\mathbf{k}$ since other factors in the integral vary much more rapidly. Then using (2), (3) and (6) in (11) we find

$$B = \frac{8C\pi^2 h\omega^2(\hbar\omega)^{1/2} n_D}{3a^3}\left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0}\right)\left[n(x_0) + \frac{1}{2}\right]$$

$$\times |\langle 0|e\mathbf{x}|\mu\rangle|^2 \int_{p_0}^\infty \frac{(p-p_0)^{1/2}}{p}\{R_p \exp(-px) - R_{-p}\}dp. \qquad \ldots\ldots(12)$$

A method for the approximate evaluation of integrals of this type in equation (12) has been given by O'Dwyer and Harper (1957) provided that the argument of the Bessel function exceeds its order and that

$$m(x) = p_0\beta^{1/2} + (x - \tfrac{1}{2}x_0)/2\beta^{1/2} \gg 1 \qquad \ldots\ldots(13)$$

where

$$\beta = \frac{1}{4S\{n(x_0)[n(x_0)+1]\}^{1/2}}. \qquad \ldots\ldots(14)$$

Using this approximation we find

$$\int_{p_0}^\infty \frac{(p-p_0)^{1/2}}{p}\{R_p \exp(-px) - R_{-p}\}dp$$

$$\simeq \frac{\exp\{\tfrac{1}{2}\beta - S[2n(x_0)+1] - p_0^2\beta\}}{2\sqrt{2}\beta^{1/4}p_0^2}$$

$$\times \left[\frac{\exp\{-p_0(x - \tfrac{1}{2}x_0)\}}{\{m(x)\}^{3/2}} - \frac{\exp(-\tfrac{1}{2}p_0 x_0)}{\{m(0) + x_0/2\beta^{1/2}\}^{3/2}}\right]. \qquad \ldots\ldots(15)$$

This expression vanishes for $x = x_0$ as required and increases with decreasing x, reaching a fixed values as x approaches zero. It is not, however, valid for $x \ll x_0$ unless $p_0 \gg S$ and it will be shown that p_0 and S are of the same order of magnitude for the traps to be considered. This means that $m \ll 1$ and since we shall be dealing with values of $x_0 \lesssim 1$ we can write to a very good approximation

---

† It is surprising at first sight that an approximate expression for the transition probabilities should give an exact balance in the equilibrium state, while the more exact expression given by Huang and Rhys and simplified by them in an approximate manner to obtain (6) does not give zero energy transfer in (9) for the case $T=T_0$. The answer lies in the equivocation " W is the ionization energy of the electron ". Properly speaking we should distinguish three such ionization energies—the optical, the lattice thermal and the electron thermal. Then in the absence of optical processes lattice ionization must balance lattice recombination and electron ionization must balance electron recombination in the equilibrium state ; the resulting conduction electron distribution would be very difficult to describe since its available energy states will vary with the lattice configuration. It is evident then that there is a certain physical inconsistency in this work since we are considering the ionization energy to be the same both for the electron–electron collisions which maintain the distribution function and for the electron–lattice processes which transfer energy to the lattice. However, the model which regards these energies as the same is not inconsistent with the approximate transition probabilities since together they yield a relation for the net energy transfer to the lattice which approaches zero as T approaches T_0.

$$m(x) \simeq \left(\frac{S}{x_0}\right)^{1/2} \left\{\frac{p_0 x_0}{2S} + x - \frac{x_0}{2}\right\} \qquad \ldots\ldots(16)$$

$$\beta \simeq x_0/4S. \qquad \ldots\ldots(17)$$

The integral in question is then given by

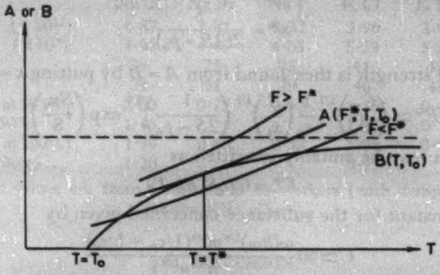
$$\int_{p_0}^{\infty} \frac{(p-p_0)^{1/2}}{p} \{R_p \exp(-px) - R_{-p}\}\, dp$$

$$\simeq \left(\frac{S}{x_0}\right)^{1/4} \frac{\exp\{-Sx + \frac{1}{4}Sx_0\}}{p_0^2 \sqrt{2}}. \qquad \ldots\ldots(18)$$

Substitution of (18) into (12) then gives an expression for B which we may use when $T \gg T_0$ and $p_0 \sim S$.

3.3. *Calculation of the Critical Field Strength*

If we plot the energy rates A and B as functions of electron temperature for a fixed lattice temperature and various values of the field strength we find the situation shown in the figure which is of the same general form as found by Fröhlich for his amorphous dielectric. For low field strengths two equilibrium electron temperatures are possible—the lower one stable and the upper one unstable. As the field strength increases the two equilibrium temperatures coalesce at a field strength which is identified with the critical field strength since above it no equilibrium electron temperature is possible. We can find this critical field strength by eliminating x from the simultaneous equations



The energy rates A and B as functions of electron temperature.

$$\left.\begin{aligned} A(F^*, x, x_0) &= B(x, x_0) \\ \frac{d}{dx} A(F^*, x, x_0) &= \frac{d}{dx} B(x, x_0) \end{aligned}\right\} \qquad \ldots\ldots(19)$$

and solving the resulting equation for F^*. A general solution of (19) would be algebraically very difficult, but it is found that $x^* \ll x_0$ for trap depths and interaction constants with which we shall be concerned. In the region $x \ll x_0$ we have from (2), (4) and (5)

$$A(F^*, x, x_0) \simeq C_A \frac{x_0 F^{*2}}{x^{5/4}} \exp(-\tfrac{1}{2} p_0 x) \qquad \ldots\ldots(20)$$

where

$$C_{\mathrm{A}} \simeq \frac{n_{\mathrm{D}}^{1/2}}{2(1/\epsilon_\infty - 1/\epsilon_0)} \left(\frac{h^2 \hbar\omega}{\pi^7 m^3}\right)^{1/4}. \qquad \ldots\ldots(21)$$

Also from (12) and (18) in the region $x \ll x_0$ we have

$$B(x, x_0) \simeq \frac{C_{\mathrm{B}}}{x_0} \left(\frac{S}{x_0}\right)^{1/4} \exp(-Sx + \tfrac{1}{4}Sx_0) \qquad \ldots\ldots(22)$$

where

$$C_{\mathrm{B}} = \frac{32\pi^3 \omega^2 (\hbar\omega)^{1/2} (2m)^{3/2}}{3\sqrt{2}a^3 h^2} \times \frac{(1/\epsilon_\infty - 1/\epsilon_0) n_{\mathrm{D}} |\langle 0|e\mathbf{x}|\mu\rangle|^2}{p_0^{\,2}}. \qquad \ldots\ldots(23)$$

From (20) and (22) we have then

$$\left.\begin{aligned} \frac{dA}{dx}(F^*, x, x_0) &= -(\tfrac{1}{2}p_0 + 5x/4)A \\ \frac{dB}{dx}(x, x_0) &= -SB. \end{aligned}\right\} \qquad \ldots\ldots(24)$$

Substituting (24) in (19) and solving we have

$$x^* = \frac{5}{2(2S - p_0)}. \qquad \ldots\ldots(25)$$

The critical field strength is then found from $A = B$ by putting $x = x^*$ which gives

$$F^* \simeq \left(\frac{C_{\mathrm{B}}}{C_{\mathrm{A}}}\right)^{1/2} \frac{1}{x_0} \left(\frac{S}{x_0}\right)^{1/8} \left(\frac{1}{2S - p_0}\right)^{5/8} \exp\left(\frac{Sx_0}{8}\right). \qquad \ldots\ldots(26)$$

This formula may be suitably rewritten as

$$F^* = Cf(S, p_0, T) \qquad \ldots\ldots(27)$$

where C is a constant for the substance concerned given by

$$C \simeq 60 \frac{\omega(\hbar\omega)^{1/8} m^{9/8} (1/\epsilon_\infty - 1/\epsilon_0)}{h^{5/4} a^{3/2}} \qquad \ldots\ldots(28)$$

and $f(S, p_0, T_0)$ is a function of the electron trap which is being considered, and the lattice temperature. It is given by

$$f(S, p_0, T_0) = \frac{|\langle 0|e\mathbf{x}|\mu\rangle| n_{\mathrm{D}}^{1/4}}{p_0 x_0} \left(\frac{S}{x_0}\right)^{1/8} \left(\frac{1}{2S - p_0}\right)^{5/8} \exp\left(\frac{Sx_0}{8}\right). \qquad \ldots\ldots(29)$$

§ 4. EVALUATION OF A CRITICAL FIELD STRENGTH FOR THE ALKALI HALIDES

The factor C given by (28) is accurately determined for the alkali halides but f (29) must be calculated from the properties of some particular type of trapping centre. We shall take the trap to be the ground state associated with an interstitial positive ion, and using the static method due to Simpson (1949) for calculating the wave function of the trapped electron we have for this wave function

$$\phi\mu = (\mu^3/\pi)^{1/2} \exp(-\mu r)$$

where $\mu = (5/\epsilon_\infty + 11/\epsilon_0)/16a_0$ with a_0 the Bohr radius. From this we readily find

$$\langle 0|e\mathbf{x}|\mu\rangle = 6\pi^{3/2} e/\mu^{5/2} \qquad \ldots\ldots(30)$$

and Simpson gives for the thermal ionization energy

$$p_0 \hbar\omega = 0{\cdot}0531(5/\epsilon_\infty + 11/\epsilon_0)^2 \mathrm{ev}. \qquad \ldots\ldots(31)$$

The interaction constant S is defined by Huang and Rhys as the energy (measured in units of $\hbar\omega$) of the electric field which is the difference between the electric fields of the initial and final states of the trapping centre. This is found on the basis of Simpson's model to be given by

$$S\hbar\omega = 1{\cdot}06\left(\frac{5}{\epsilon_\infty} + \frac{11}{\epsilon_0}\right)\left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_0}\right)\text{ev.} \qquad \ldots\ldots(32)$$

It is then only required to know n_D the number of trapping centres. This must be estimated but is fortunately not a critical quantity since it occurs only to the one quarter power. For the alkali halides which we shall consider, taking $n_D{}^{1/4} = 2{\cdot}5 \times 10^4$ means a relative defect density of 1 in 10^5, which seems quite reasonable for pure single crystals. In the table we have set out the results of some calculations on the critical field strengths for the chlorides, bromides and iodides of sodium and potassium at 50°c and 100°c. We have also given experimental values of breakdown strength at these temperatures where they are available.

Calculated Critical Field Strengths and Experimental Breakdown Strengths (mv cm^{-1}) for some Alkali Halides

		Na Cl	Na Br	Na I	K Cl	K Br	K I
ω sec^{-1}	$(\times 10^{13})$	4·87	3·81	3·32	3·96	3·06	2·52
u cm^{-1}	$(\times 10^{7})$	4·96	4·45	4·03	5·59	5·30	4·92
p_0		24	25	25	41	48	49
S		31	31	29	46	48	44
F^* (Eqn (27) at 50°c)		2·00	0·75	0·43	1·36	0·44	0·24
F^* (Expt at 50°c)		1·45			1·15	0·73 ?	
F^* (Eqn (27) at 100°c)		1·59	0·68	0·39	1·01	0·34	0·15
F^* (Expt at 100°c)		1·30			0·95		

The experimental values are from Cooper and co-workers (with temperature uncertain for the KBr values).

§ 5. Discussion

Since the calculated critical field strengths are of the same order of magnitude, show the same variation from substance to substance, and have almost the same temperature variation as the experimental breakdown strengths, it seems reasonable to postulate that the mechanism described by the theory is the breakdown mechanism in the high temperature region. However, there are some objections to this which must be mentioned.

In the first place the electron–phonon interaction has been ignored as a process for transferring electron energy to the lattice. This is certainly not permissible for low electron temperatures, but for electron temperatures T such that $kT \gtrsim 10\hbar\omega$ with which we have been dealing, a comparison of (22) of this work with (24) of O'Dwyer (1954) shows that transfer of energy by radiationless recombination exceeds that by inelastic collisions with phonons by a factor of about 10^3.

Secondly these high electronic temperatures must be attained and it has been shown (cf. Fröhlich 1952) that a Maxwellian distribution of conduction electrons interacting with the lattice only by electron–phonon interaction will, under the application of a strong field, reach a temperature of about twice the lattice temperature, and that the critical field strength is an increasing function of lattice temperature. Since the present model deals with deep traps the number of conduction electrons will vary only slightly with T when $T \lesssim 2T_0$ and the calculations of Fröhlich will apply. If then his critical field strength is exceeded electron–phonon interaction is no longer capable of producing an equilibrium electron temperature, and the present work should be viewed as an attempt to find out whether another mechanism (radiationless recombination) can re-stabilize the electron distribution at a higher temperature. It follows of course that if, in any region, the critical field of the present work is the breakdown field, then in that region it must be greater than the critical field of the type first calculated in Fröhlich (1952).

The third and most serious objection is that n_D is a more or less disposable constant, and while it is true that it cannot vary by many orders of magnitude (and hence $F^* \propto n_D^{1/4}$ varies much less), nevertheless n_D must have some temperature dependence and it seems that this must be invoked to explain the non-applicability of the present ideas to the region of low temperature breakdown.

Provided that other factors do not intervene an increase in the number of trapping centres should raise the critical field. It would be interesting to see whether the addition of divalent positive impurities would raise the high temperature breakdown strength of alkali halides.

It seems not to matter greatly that we have used interstitial positive ions as trapping centres, since the characteristics of any type of trap should vary in a similar manner from crystal to crystal.

REFERENCES

ALGER, R. S., and VON HIPPEL, A., 1949, *Phys. Rev.*, **76**, 127.
AUSTEN, A. E. W., and WHITEHEAD, S., 1940, *Proc. Roy. Soc.* A, **176**, 33.
CALDERWOOD, J. H., and COOPER, R., 1953, *Proc. Phys. Soc.* B, **66**, 73.
CALLEN, H. B., 1949, *Phys. Rev.*, **76**, 1394.
COOPER, R., and GROSSART, D. T., 1953, *Proc. Phys. Soc.* B, **66**, 716.
FRÖHLICH, H., 1937, *Proc. Roy. Soc.* A, **160**, 230 ; 1939, *Ibid.*, **172**, 94 ; 1947, *Ibid.*, **188**, 521 ; 1952, *E.R.A. Rep. L/T277.*
FRÖHLICH, H., and PARANJAPE, B. V., 1956, *Proc. Phys. Soc.* B, **69**, 21.
HUANG, K., and RHYS, A., 1950, *Proc. Roy. Soc.* A, **204**, 1950.
O'DWYER, J. J., 1954, *Aust. J. Phys.*, **7**, 36.
O'DWYER, J. J., and HARPER, P. G., 1957, *Phys. Rev.*, **105**, 399.
SEITZ, F., 1940, *Modern Theory of Solids* (New York : McGraw Hill).
SIMPSON, J. H., 1949, *Proc. Roy. Soc.* A, **197**, 269.

Formulae for Dielectric Constant of Mixtures

By J. A. REYNOLDS† and J. M. HOUGH
Department of Physics, The University, Hull

MS. received 28th January 1957

Abstract. The major formulae suggested for the dielectric constant of a mixture of two components (at least one of which has the form of particles containing many molecules) are shown to be all related to another. The assumptions made in deriving the differing formulae are examined. The type of formula suggested by Lichtenecker does not fit in this generalization and it is suggested that this is due to incorrect theoretical assumptions.

§ 1. INTRODUCTION

THE dielectric constant of a mixture is a function of the volume fraction, shape and dielectric constant of each component. Theoretical calculations are most easily made for systems where the components are arranged in a regular array but normal mixtures show a random arrangement of the components. Over the last fifty or sixty years many formulae have been suggested for such mixtures ; the method of derivation has varied greatly and it is not always obvious what relation exists between the formulae. It is the purpose of this paper to show that practically all the published formulae are in fact special cases of a general formula.

In order to avoid complications the treatment is restricted to a mixture of only two components both of which are isotropic and randomly distributed in space. The extension to more complicated cases is fairly obvious and has been discussed in the literature (e.g. Bruggeman 1935). The case of regular arrays has recently been discussed by Kharadly and Jackson 1953) ; it will not be considered here.

§ 2. DERIVATION OF THE GENERAL FORMULA

Let the two substances have dielectric constants ϵ_1 and ϵ_2, and occupy fractions of the total volume of the mixture δ_1 and δ_2. Then

$$\delta_1+\delta_2=1. \qquad \ldots\ldots(1)$$

Consider the electric displacement D, in the mixture ; its value averaged over all the volume V is $\bar{D}=\int_V D dV/V$. Over the component with dielectric constant ϵ_1 and volume V_1, the average value is $\bar{D_1}=\int_{V_1} D\,dV/V_1$ and similarly over the second component. Therefore

$$\bar{D}=\frac{1}{V}\int_V D\,dv=\frac{1}{V}\left[\int_{V_1} D\,dV+\int_{V_2} D\,dV\right]=\delta_1\bar{D}_1+\delta_2\bar{D}_2. \qquad \ldots\ldots(2)$$

In the same way the average field $\bar{E}$ is given by

$$\bar{E}=\delta_1\bar{E}_1+\delta_2\bar{E}_2. \qquad \ldots\ldots(3)$$

---

† Now at the Atomic Weapons Research Establishment, Aldermaston, Berks.

Following Wiener (1912), it may be assumed that the dielectric constant of the mixture is given by

$$\epsilon = \bar{D}/\bar{E}$$

and that for each component $\bar{D}_1 = \epsilon_1 \bar{E}_1$ and $\bar{D}_2 = \epsilon_2 \bar{E}_2$. From equations (2) and (3)

$$\epsilon = \epsilon_1 \delta_1 f_1 + \epsilon_2 \delta_2 f_2 \qquad \ldots\ldots(4)$$

$$1 = \delta_1 f_1 + \delta_2 f_2 \qquad \ldots\ldots(5)$$

where $f_1 = \bar{E}_1/\bar{E}$ and $f_2 = \bar{E}_2/\bar{E}$. Rearranging these equations yields

$$\epsilon = \epsilon_2 + (\epsilon_1 - \epsilon_2)\delta_1 f_1 \qquad \ldots\ldots(6)$$

or

$$(\epsilon - \epsilon_1)\delta_1 f_1 + (\epsilon - \epsilon_2)\delta_2 f_2 = 0. \qquad \ldots\ldots(7)$$

Equations (6) and (7) are the general forms of the formulae for the dielectric constants of mixtures. Theoretically the two equations are equivalent but when approximations have to be made for the values of the f's, that is no longer true. It would seem appropriate to use equation (6) for the case of particles dispersed in a continuous medium and equation (7) for the case where the particle size of the two components was of the same order of magnitude. This second case was first treated extensively by Bruggeman (1935); later, Odelevskiĭ (1951) named such a system a statistical mixture.

§ 3. DETERMINATION OF THE FIELD RATIO f

Before equations (6) and (7) can be used it is necessary to know the value of the field ratio f_1 (and f_2 where required). Unfortunately, it is only possible to calculate it exactly for the case of parallel slabs or the case of infinitely diluted dispersions of particles of ellipsoidal shape; in all other cases it is necessary to make an approximate calculation. Most of the published formulae differ from one another in the approximation considered appropriate.

The average value of the field within a particle depends upon its shape, dielectric constant and the average field and dielectric constant of the medium immediately outside it. The normal simplification for the external conditions is to assume that the particle is immersed in a homogeneous medium of dielectric constant ϵ^* with an electric field E. In the literature four different values of ϵ^* have been suggested; (1) $\epsilon^* = \epsilon_2$, (2) $\epsilon^* = \epsilon$, (3) an integration method due to Bruggeman (1935) and (4) the experimental determination of ϵ^* (de Loor 1954).

For a very dilute dispersion the interaction of the particles may be ignored and then $\epsilon^* = \epsilon_2$ is the appropriate approximation to use. Many of the early formulae used this approximation e.g. Lorentz (1880), Wagner (1914a, b) and more recently Ollendorf (1931) and Sillars (1937). Formulae based on this approximation will be satisfactory only for low concentrations of dispersed particles and for higher concentrations allowance must be made for the interaction of the dispersed particles. The simplest method is to make $\epsilon^* = \epsilon$ i.e. to the dielectric constant of the mixture. This was done (not explicitly) by Bruggeman (1935) and explicitly by Böttcher (1945); more recently it has been used by Polder and van Santen (1946) and Niesel (1952). The integration method will be discussed in a later section (§ 5) and the fourth method is chiefly of value to see how satisfactory the other approximations are.

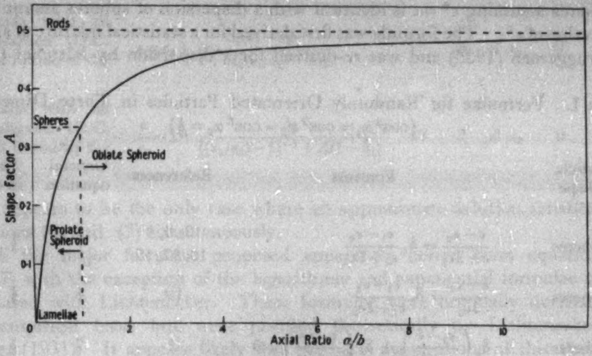
These approximations reduce the problem to that of finding the induced field when a particle is placed in a homogeneous field. Once again this problem is only soluble for the specially simple case of ellipsoids and in practice the only ellipsoids actually considered are the spheroids. Two limiting cases of the spheroids are infinitely long rods and infinitely thin circular discs; these two cases provide good approximations for the rod and lamallae types of particles often met with in practice. De Loor (1954) has given an experimental method of deducing the most appropriate eccentricity to use as a substitute for the actual particle considered.

§ 4. FIELD FACTOR *f* FOR A SPHEROID

The induced field inside a spheroid in a homogeneous field has been derived in the standard text-books of electricity (e.g. Stratton 1941) and leads to the expression

$$f_1 = \sum_{i=1}^{3} \frac{\cos^2 \alpha_i}{1 + A_i[(\epsilon_1/\epsilon^*) - 1]} \qquad \ldots\ldots(8)$$

where α_i are the angles made by the spheroid axes and the applied field and the A_i depend on the axial ratios of the spheroid. $(A_1 + A_2 + A_3 = 1)$. For a spheroid



Value of *A* for spheroids.

$A_2 = A_3 = A$ and $A_1 = 1 - 2A$ and a graph of A as a function of the axial ratio is given in the figure. For a random orientation of the spheroids

$$\cos^2 \alpha_1 = \cos^2 \alpha_2 = \cos^2 \alpha_3 = \tfrac{1}{3}$$

and for the special case of long particles aligned with the long axes parallel but otherwise random, $\cos^2 \alpha_1 = \cos^2 \alpha_2 = \frac{1}{2}$; $\cos^2 \alpha_3 = 0$.

§ 5. INTEGRATION METHOD FOR ϵ^*

Bruggeman (1935) devised an alternative method of allowing for the interaction of the particles. He used a formula applicable to very dilute suspensions (which used $\epsilon^* = \epsilon_2$) to find the increase in dielectric constant of the mixture caused by making a small addition of one component; integration between limits yielded the dielectric constant of the mixture. For statistical mixtures this in fact merely amounted to the assumption $\epsilon^* = \epsilon$. For dispersions Bruggeman treated lamellae and spheres and the calculation was extended to rods by Niesel (1952). Integration for the general case of randomly orientated spheroids gives

$$\epsilon=\epsilon_1-(\epsilon_1-\epsilon_2)(1-\delta_1)\left(\frac{\epsilon}{\epsilon_2}\right)^{\alpha}\left(\frac{\epsilon_1\beta+\epsilon}{\epsilon_1\beta+\epsilon_2}\right)^{\gamma} \qquad \ldots\ldots(9)$$

where

$$\alpha=\frac{3A(1-2A)}{2-3A}\ ;\ \beta=\frac{2-3A}{1+3A}\ ;\ \gamma=\frac{2(3A-1)}{(1+3A)(2-3A)}.$$

This approach does not directly yield a value of ϵ^* but it may be shown that the effective value of ϵ^* lies between ϵ and ϵ_2. This type of behaviour has in fact been found experimentally by de Loor (1954).

§ 6. PARTICULAR FORMULAE

Most of the formulae proposed by different workers are summarized in tables 1 and 2. In table 1, formulae for particles randomly orientated in three dimensions are summarized whereas in table 2 the particles are orientated in one or more directions.

A point which is of some interest is that the formulae for statistical mixture of spheres assuming $\epsilon^*=\epsilon$ is identical with a dispersion of spheres assuming the same value of ϵ^*. The formula was first derived for a statistical mixture of spheres by Bruggeman (1935) and was re-derived for a dispersion by Böttcher (1945).

Table 1. Formulae for Randomly Orientated Particles in Three Dimensions ($\cos^2\alpha_1=\cos^2\alpha_2=\cos^2\alpha_3=\frac{1}{3}$)

Particle Shape	Formula	References	General equation	Factors in eqn 8 A	ϵ^*
Spheres	$\frac{\epsilon-\epsilon_2}{\epsilon+2\epsilon_2}=\delta_1\frac{\epsilon_1-\epsilon_2}{\epsilon_1+2\epsilon_2}$	3, 8, 9 10, 12, 15	7	$\frac{1}{3}$	ϵ_2
Spheres	$\frac{\epsilon-\epsilon_2}{3\epsilon_2}=\delta_1\frac{\epsilon_1-\epsilon_2}{\epsilon_1+2\epsilon_2}$	19	6	$\frac{1}{3}$	ϵ_2
Spheres	$\delta_1\left(\frac{\epsilon_1-\epsilon}{\epsilon_1+2\epsilon}\right)+\delta_2\left(\frac{\epsilon_2-\epsilon}{\epsilon_2+2\epsilon}\right)=0$	2	7	$\frac{1}{3}$	ϵ
Spheres	$\frac{\epsilon-\epsilon_2}{3\epsilon}=\delta_1\frac{\epsilon_1-\epsilon_2}{\epsilon_1+2\epsilon}$	1	6	$\frac{1}{3}$	ϵ
Spheroids	$\epsilon=\epsilon_2+\frac{\delta_1}{3(1-\delta_1)}\sum_{i=1}^{3}\frac{\epsilon_1-\epsilon}{1+A_i\left(\frac{\epsilon_1}{\epsilon_2}-1\right)}$	4	7	A	ϵ_2
Spheroids	$\epsilon=\epsilon_2+\frac{\delta_1}{3}\sum_{i=1}^{3}\frac{\epsilon_1-\epsilon_2}{1+A_i\left(\frac{\epsilon_1}{\epsilon}-1\right)}$	14	6	A	ϵ
Lamellae	$\epsilon^2=\frac{2(\epsilon_1\delta_1+\epsilon_2\delta_2)-\epsilon}{\epsilon_1/\delta_1+\epsilon_2/\delta_2}$	2	7	0	ϵ
Rods	$5\epsilon^3+(5\epsilon_p{}'-4\epsilon_p)\epsilon^2$ $-(\delta_1\epsilon_1{}^2+4\epsilon_1\epsilon_2+\delta_2\epsilon_2{}^2)-\epsilon_1\epsilon_2\epsilon_p=0$ where $1/\epsilon_p=\delta_1/\epsilon_1+\delta_2/\epsilon_2$ $1/\epsilon_p{}'=\delta_1/\epsilon_2+\delta_2/\epsilon_1$	11	6	$\frac{1}{2}$	ϵ

Table 2. Formulae for Particles with Some Degree of Orientation

Particle Shape	Formula	Ref.	Gen. eqn	Factors in eqn 8			
				A	ϵ^*	$=\begin{matrix}\cos\alpha_1\\ \cos\alpha_2\end{matrix}$	$\cos\alpha_3$
Parallel cylinders	$\frac{\epsilon-\epsilon_2}{\epsilon+\epsilon_2}=\delta_1\frac{\epsilon_1-\epsilon_2}{\epsilon_1+\epsilon_2}$	12, 15	7	$\frac{1}{2}$	ϵ_2	$\frac{1}{2}$	0
Parallel cylinders	$\delta_1\frac{\epsilon_1-\epsilon}{\epsilon_1+\epsilon}+\delta_2\frac{\epsilon_2-\epsilon}{\epsilon_2+\epsilon}=0$	2	7	$\frac{1}{2}$	ϵ	$\frac{1}{2}$	0
Parallel lamellae (with two axes randomly orientated)	$\epsilon^2=\frac{\epsilon_1\delta_1+\epsilon_2\delta_2}{\epsilon_1/\delta_1+\epsilon_2/\delta_2}$	2	7	0	ϵ	$\frac{1}{2}$	0
Lamellae with all axes aligned (D.C. perpendicular to the lamellae planes)	$\frac{1}{\epsilon}=\frac{\delta_1}{\epsilon_1}+\frac{\delta_2}{\epsilon_2}$	20	7	0	ϵ	0	1
Lamellae with all axes aligned (D.C. parallel to the lamellae planes)	$\epsilon=\epsilon_1\delta_1+\epsilon_2\delta_2$	16, 20	7	0	ϵ	1	0
Spheroids with all axes aligned (D.C. parallel to one of the axes)	$\epsilon=\epsilon_2+\frac{\delta_1(\epsilon_1-\epsilon_2)}{1+A[(\epsilon_1/\epsilon_2)-1)]}$	17	6	A	ϵ_2	0	1
Spheroids with all axes aligned (D.C. parallel to one of the axes)	$\frac{\epsilon}{\epsilon_2}=1+\frac{\delta_1}{[(\epsilon_1/\epsilon_2)-1]^{-1}+A(1-\delta_1)}$	13	7	A	ϵ_2	0	1

This appears to be the only case where an approximate solution satisfied both equations (6) and (7) simultaneously.

All the major formulae proposed apparently derive from equations (6) and (7) with the exception of the logarithmic and exponential formulae usually associated with Lichtenecker. These formulae were originally derived on a semi-empirical basis but were justified theoretically by Lichtenecker and Rother (1931). It appears likely that the basic assumptions of this theory are incorrect (see Appendix).

§ 7. CONCLUSIONS

The general theory of the dielectric constant of mixtures seems to be adequately represented by equations (6) and (7) but, in practice, it is not possible to use these formulae directly owing to the difficulty of calculating the field ratios for the general case. All theoretical work on random mixtures reduces to the determination of suitable approximations for the field ratios; it has been shown that the principle formulae used are all special cases of these fundamental equations.

In making approximations, it is useful to distinguish between statistical mixtures and dispersions and to use equations (7) and (6) respectively. By using equation (8) the field ratios can be calculated for ellipsoidal particles dispersed in a homogeneous medium of dielectric constant ϵ^*. Three methods of approximating to the actual medium by choosing an appropriate value of ϵ^* have been used and the choice of the most suitable approximation requires an experimental investigation. It is hoped shortly to publish an account of some experimental observations on dispersions designed to compare the validity of these assumptions.

ACKNOWLEDGMENTS

These investigations were performed in the Physics Department of the University of Hull and the authors are indebted to Professor L. S. Palmer for suggesting the problem and for his help and encouragement.

One of us (J. A. R.) is indebted to the University of Hull for a Research Scholarship.

APPENDIX

A theoretical basis of the Lichtenecker type of formula was given by Lichtenecker and Rother (1931) and is based on an assumption of repeated mixing, namely that the final dielectric constant of a mixture is independent of the method of preparation of the mixture.

Consider a mixture of two components of dielectric constants ϵ_1 and ϵ_2 with a volume fraction δ_1 of the first component; then the dielectric constant of the mixture can be expressed as

$$\epsilon = f(\epsilon_1, \epsilon_2, \delta_1). \qquad \ldots\ldots(a)$$

Consider now two other mixtures A and B containing volume fractions δ_A, δ_B of the first component. Then the dielectric constants of the mixtures are given by

$$\epsilon_A = f(\epsilon_1, \epsilon_2, \delta_A); \; \epsilon_B = f(\epsilon_1, \epsilon_2, \delta_B). \qquad \ldots\ldots(b)$$

A mixture of A and B can be made with the volume fraction δ_M of A such that the composition of the final mixture is identical with that of the original mixture. Then an equation similar to equation (*a*) can be set up

$$\epsilon_M = F(\epsilon_A, \epsilon_B, \delta_M). \qquad \ldots\ldots(c)$$

Rother assumed

$$f(\epsilon_1, \epsilon_2, \delta_1) = F(\epsilon_1, \epsilon_2, \delta_1) \qquad \ldots\ldots(d)$$

$$\epsilon = \epsilon_M. \qquad \ldots\ldots(e)$$

It would appear that there is no reason for both (*d*) and (*e*) to hold simultaneously. If the two stage mixing is done without change of the state of the individual components, there appears to be no justification for (*d*) as there are no 'particles' of A and B. If, on the other hand, 'particles' of A and B are formed, by some process before the second mixing, there seems no justification for (*e*).

Thus the theoretical basis of the Lichtenecker type formula is not valid and this is the reason why it does not satisfy the general equations derived here.

REFERENCES

1. Böttcher, C. J. F., 1945, *Rec. Trav. Chim. Pays-Bas*, **64**, 47.
2. Bruggeman, D. A. G., 1935, *Ann. Phys., Lpz.* (5), **24**, 636.
3. Clausius, R., 1879, *Die Moch. Warme Theorie*, II (Braunschweig).
4. Fricke, H., 1924, *Phys. Rev.*, **24**, 575 ; 1931, *Physics* **1**, 106.
5. Kharadly, M. M. Z., and Jackson, W., 1953, *Proc. Inst Elect. Engrs*, **100**, 199.
6. Lichtenecker, K., and Rother, K., 1931, *Phys. Z.*, **32**, 255.
7. de Loor, G. P., 1954, *Appl. Sci. Res.* B, **3**, 479.
8. Lorentz, H. A., 1880, *Ann. Phys., Lpz.*, **9**, 641.

9. Lorenz, L., 1880, *Ann. Phys., Lpz.*, **11**, 70.
10. Mosotti, O. F., 1850, *Mem. de mathem et di fisica in Modena* II, **24**, 49.
11. Niesel, W., 1952, *Ann. Phys., Lpz.*, **10**, 336.
12. Odelevskiĭ, V. I., 1951, *Zhurnal Technicheskai*, **6**, 667, 678.
13. Ollendorf, F., 1931, *Arch. Elektrochem.*, **25**, 436.
14. Polder, D., and van Santen, J. M., 1946, *Physica*, **12**, 257.
15. Rayleigh, Lord, 1892, *Phil. Mag.*, **34**, 481.
16. Silberstein, L., 1895, *Ann. Phys., Lpz.*, **56**, 661.
17. Sillars, R. W., 1937, *J. Inst. Elect. Engrs.*, **80**, 378.
18. Stratton, J. A., 1941, *Electromagnetic Theory* (*New York : McGraw-Hill*).
19. Wagner, K. W., 1914 a, *Arch. Elektrochem.*, **2**, 371 ; 1914 b, *Ibid.*, **3**, 100.
20. Wiener, O., 1912, *Abh. sächs. Ges.* (*Akad.*) *Wiss.*, **32**, 509.

Absorption and Dispersion of Indium Antimonide

By T. S. MOSS, S. D. SMITH and T. D. F. HAWKINS
Radio Department, Royal Aircraft Establishment, Farnborough, Hants.

MS. received 18*th April* 1957

Abstract. Transmission measurements have been used to determine absorption coefficients over the wavelength range 1·5–7·5 μ, and interferometric determinations made of the refractive index from 7–20 μ. It is shown that the following four factors contribute significantly to the dispersion: (i) short-wave absorption band centred in the visible region, (ii) tail absorption band, (iii) free carrier absorption, (iv) Restrahlen band. The 'characteristic' value of refractive index occurs at approximately 10 μ, where it is 3·96.

A more accurate calculation is made of the radiation lifetime. Using these data, $\tau_R = 0{\cdot}79\,\mu$sec.

§ 1. Introduction

Knowledge of the optical properties of InSb forms an important part of the fundamental study of this interesting semiconductor, and is required to enable theoretical estimates of photoelectric behaviour and radiation lifetime to be made. The optical properties are of particular interest in themselves, since as a result of the absorption edge lying at very long wavelengths, the intrinsic carrier concentration being high and the effective mass of the free electrons being very small, there is considerable dispersion throughout the whole infra-red region of the spectrum.

Earlier estimates of the refractive index of InSb have been given by (i) Avery *et al.* (1954) who from polarization reflectivity measurements at short wavelengths find $n = 3{\cdot}7$ for impure specimens, (ii) Oswald and Schade (1954) who find $n = 4{\cdot}1$ from the magnitude of reflectivity between 7 and 16 μ, (iii) Briggs *et al.* 1954, by a prism method (the value is quoted by Fan (1956) at an indeterminate wavelength in the highly dispersive region), (iv) Breckenridge *et al.* (1954), (v) Yoshinaga and Oetjen (1956) from the magnitude of reflectivity for wavelengths beyond 20 μ. In the present work the refractive index has been measured interferometrically between 7 and 20 μ and the dispersion analysed in detail.

A preliminary account of our absorption measurements was given by Moss, Hawkins and Smith at the Physical Society Conference on Semiconductors in April 1956; the only other high absorption measurements on pure material known are unpublished work of Gobeli quoted by Fan (1956). Our results are compared with theoretical estimates by Kane (1957).

§ 2. Experimental Details

All measurements were made using a Leiss double monochromator. For the transmission work, LiF and CaF_2 prisms were used. For the reflection interference measurements CaF_2, NaCl and KBr prisms were used. At short wavelengths the detector was a PbS cell, elsewhere it was a thermopile.

The material was single crystal, intrinsic InSb. The refractive index measurements were made on self-supporting specimens of thickness ranging from 16 μ to approximately 100 μ, the upper limit being determined by the energy limited resolution of the spectrometer.

As InSb is mechanically weak (much more so than Ge for example) preparation of the thinner specimens was difficult. One technique which proved effective was to stick a slice of material on a thin glass slide with molten shellac (after first polishing the underside). The upper surface was then ground down and polished until interference measurements in reflection showed that the surfaces were parallel and the thickness suitable. A hole was then dissolved in the glass backing opposite the centre of the specimen with hydrogen fluoride, the shellac protecting the specimen. Finally, the shellac inside this hole was dissolved in hot alcohol, leaving the specimen unbacked and conveniently mounted.

For the thinnest samples used in the transmission work, specimens were firmly stuck down on periclase plates by a thin film of thermo-setting resin (Araldite) before polishing down to the required thinness. Independent measurements established the wavelengths to be used to avoid inaccuracies due to absorption bands in the Araldite.

In the absorption measurements, the transmission was determined relative to that at wavelengths beyond the edge where the known absorption coefficient ($\sim$10 cm^{-1}) would give insignificant loss in the thin specimens used. This procedure eliminated both reflection losses and possible losses due to changing the geometry of the optical system on inserting the specimen in the beam. Slight absorption in the MgO plates at the longest wavelengths was allowed for. All specimen thicknesses were determined interferometrically. Scatter in the double monochromator was not serious for absorptions of 2000 : 1 in the highly absorbing short wavelength region. In this high absorption region optical attenuators were necessary to ensure that the range of linearity of the electronic system was not exceeded. Absorption coefficients should be accurate to within ± 15%.

§ 3. ABSORPTION RESULTS

Figure 1 shows the absorption coefficient K. It rises to $1{\cdot}6 \times 10^4$ at the short wavelength limit of measurement. Empirically the absorption constant is accurately proportional to $1/\lambda$ for wavelengths below 4 μ, i.e. at wavelengths well below the edge.

Bardeen, Blatt and Hall (1956) have derived theoretical expressions for the shape of the absorption edge for direct and indirect transitions assuming parabolic energy bands. However, a complete analysis of the band structure of InSb has recently been made by Kane (1957) who finds that the bottom of the conduction band is far from parabolic in shape, so that a simple comparison with their expressions would not be appropriate. Kane has made detailed calculations of the expected absorption spectrum for InSb using his computed energy bands on the basis of direct transitions. His results are shown as a broken line in figure 1. The extent of the agreement lends confirmation to Kane's picture of the energy bands, with valence band maximum and conduction band minimum being vertically above one another on the (E, k) diagram.

§ 4. REFRACTIVE INDEX

A typical set of reflection fringes is shown in figure 2, and a plot of the positions of fringe minima against fringe number in figure 3. It will be seen that over a

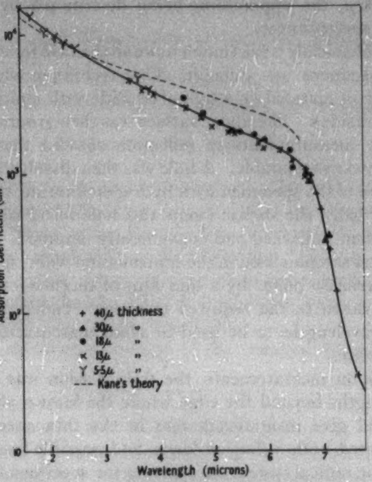


Figure 1.

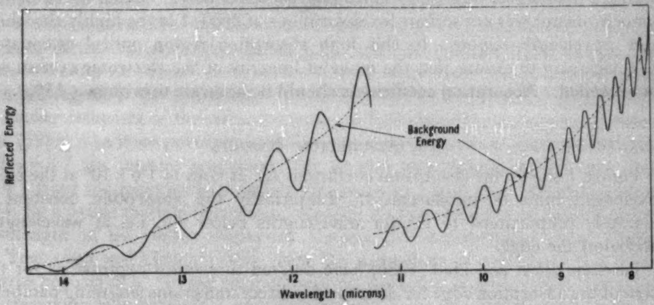


Figure 2.

wide range the fringes are equally spaced to a high degree of accuracy. This does *not* however, prove that the material is non-dispersive in this region, as the following analysis shows:

The conditions for two successive transmission maxima in a layer of thickness d at wavelengths λ_1 and λ_2 are

$$2n_1d = N\lambda_1, \qquad 2n_2d = (N+1)\lambda_2, \qquad \ldots\ldots(1)$$

where N is the fringe number. Suppose

$$n = n_0 + \beta\lambda \text{ where } \beta \text{ is a constant.} \qquad \ldots\ldots(2)$$

Then the above equations give

$$2n_0d(1/\lambda_2 - 1/\lambda_1) = 1. \qquad \ldots\ldots(3)$$

Thus β is eliminated, and can therefore have any value, i.e. a study of fringe separation *cannot* evaluate a linear dispersion. To do this it is necessary to assign the correct fringe numbers N and then calculate the refractive index at each fringe.

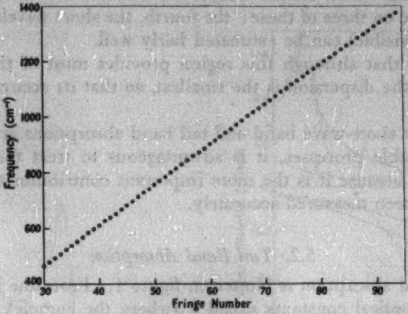


Figure 3.

In the present work the method used was to perform fringe measurements on layers sufficiently thin to determine the fringe number unambiguously. A relative dispersion curve was then calculated for the wavelength range 8–17 μ, and this curve used to identify the fringes on the next thicker specimen. From this a more accurate dispersion curve was obtained which was used to assign fringe numbers on a still thicker specimen and so on. Finally, the thickness of the thickest specimen was measured and an accurate, absolute, dispersion curve obtained.

The thickness of the actual part of the specimen used for the fringe results was measured between a pair of accurately spherical anvils driven by a Fabry–Perot interferometer. The accuracy of this determination was not worse than to $\pm\frac{1}{2}\%$.

The complete dispersion curve is shown in figure 4. The broken line is the calculated dispersion as discussed in the next section.

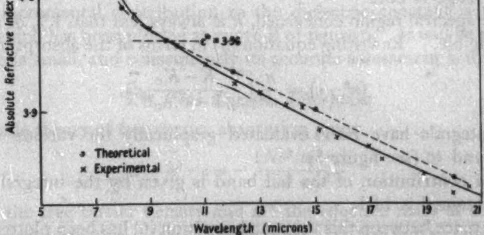


Figure 4.

§ 5. DISCUSSION OF DISPERSION DATA

5.1. *Origin of Dispersion*

As stated earlier, significant contributions are made to the dispersion by four absorption regions which can be treated separately. Direct experimental data are now available on three of these; the fourth, the short wavelength absorption (centred in the visible) can be estimated fairly well.

It transpires that although this region provides most of the absorption its contribution to the dispersion is the smallest, so that its accurate assessment is not essential.

Although the short-wave band and tail band absorptions are not necessarily physically separable processes, it is advantageous to treat the tail absorption band separately because it is the more important contributor to the dispersion and it has now been measured accurately.

5.2. *Tail Band Absorption*

The tail band absorption is plotted in figure 1. From the theoretical inter-relation of the optical constants n and k (where the complex index is $n-ik$), we have (see Moss 1952)

$$(n^2-k^2-\epsilon)_c=\frac{2}{\pi}\int_0^\infty \frac{2nk\nu-(2nk\nu)_c}{\nu^2-\nu_c{}^2}\,d\nu \qquad \ldots\ldots(4)$$

which gives the real part of the dielectric constant at a specific frequency ν_c in terms of the imaginary part at all frequencies. Here ϵ is the constant value which n^2 would have at long wavelengths if there were no additional dispersion due to free carrier and Restrahlen absorption. The contribution from the tail band may be written

$$(n^2-k^2-\epsilon)_c=\frac{4n}{\pi}\int_0^{\nu_{1\cdot 5}} \frac{k\nu-(k\nu)_c}{\nu^2-\nu_c{}^2}\,d\nu \qquad \ldots\ldots(5)$$

where $\nu_{1\cdot 5}$ is the frequency at $1{\cdot}5\,\mu$.

It is assumed that, as we are calculating small increments in refractive index, n can be treated as a constant on the right-hand side of equation (5) over the restricted waveband involved. Figure 4 shows that the variation in n is only a few per cent, and the effect of this assumption will be to introduce this order of error into the assessment of the integral. As the important part of the integral is situated around the absorption edge, an average value of refractive index in this wavelength range is used.

Over the spectral region concerned, k is always less than 1% of n, so that k^2 is quite negligible. Rewriting equation (5) in terms of the absorption coefficient

$$(n^2-\epsilon)_c=\frac{n}{\pi^2}\int_{1\cdot 5}^\infty \frac{K-K_c}{1-\lambda^2/\lambda_c{}^2}\,d\lambda. \qquad \ldots\ldots(6)$$

These integrals have been evaluated graphically for various values of λ_c between 5μ and ∞ (see figure 5).

The total contribution of the tail band is given by the integral for $\lambda_c=\infty$. The value obtained is 1·01.

The difference between this value and equation (6) has been plotted as curve A of figure 6.

It should be noted that a constant level of absorption (such as exists beyond the edge in InSb) does not contribute to the dispersion.

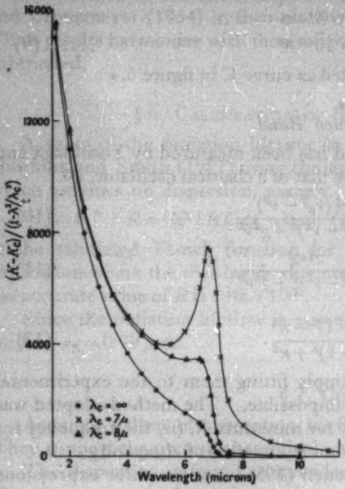


Figure 5.

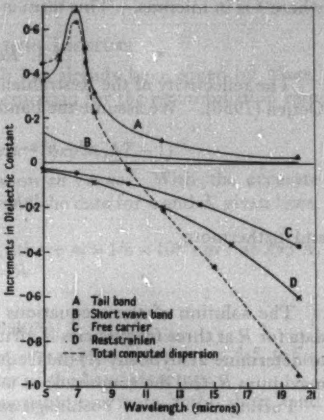


Figure 6.

5.3. *Short Wave Band Absorption*

At wavelengths not much beyond the absorption edge the dielectric constant is determined by the absorption in the short-wave and tail bands. The total contribution of the former is thus the difference between ϵ and the term from equation (6), i.e. $\epsilon-1{\cdot}01$.

The short-wave band will be assumed to be of classical oscillator form, and by analogy with germanium (which has a similar refractive index) will be assumed to be centred at $0{\cdot}5\mu$. At the relatively long wavelengths considered here, this contribution thus becomes

$$n^2-1=\frac{\epsilon-1{\cdot}01-1}{1-(0{\cdot}5)^2/\lambda^2}=(\epsilon-2{\cdot}01)(1+\tfrac{1}{4}\lambda^2). \qquad \ldots\ldots(7)$$

Hence the incremental contribution to the dielectric constant is $(\epsilon-2{\cdot}01)/4\lambda^2$ i.e. $3{\cdot}5\,\lambda^{-2}$ which has been plotted as curve B of figure 6. It will be seen that this contribution is small, and consequently its accurate assessment is unimportant.

5.4. *Free Carrier Absorption*

The Drude theory of free carrier absorption gives

$$n^2-k^2-\epsilon=-\frac{(Ne^2/m^*\epsilon_0)}{\omega^2+\gamma^2} \qquad \ldots\ldots(8)$$

where N is the free carrier density and m^* the effective mass of the carriers, ω the angular frequency and γ the damping term. In the region concerned $\omega \gg \gamma$ and k^2 is negligible as before. Hence $n^2-\epsilon=-Ne^2/m^*\epsilon_0\omega^2$. In view of the mass dependence of this expression, only the contribution of the electrons will be considered. Using the following parameters for intrinsic InSb at 295°K, $N=1\cdot6\times10^{22}\,\mathrm{m}^{-3}$ (Hrostowski *et al.* 1955, Welker and Weiss 1956), $m^*/m=0\cdot015$ (Burstein, Picus and Gebbie 1956)†. We obtain

$$n^2-\epsilon=9\cdot2\times10^{-4}\lambda^2 \qquad \ldots\ldots(9)$$

where λ is in microns. This term is plotted as curve C in figure 6.

5.5. *Restrahlen Band*

The reflectivity of the Restrahlen band has been measured by Yoshinaga and Oetjen (1956). We assume the band to be that of a classical oscillator, i.e.

$$n^2-k^2-\epsilon=\frac{A(\nu_0^2-\nu^2)}{(\nu_0^2-\nu^2)^2+g^2\nu^2}, \qquad \ldots\ldots(10)$$

$$2nk=\frac{A\nu g}{(\nu_0^2-\nu^2)^2+g^2\nu^2} \qquad \ldots\ldots(11)$$

and furthermore

$$R=\frac{(n-1)^2+k^2}{(n+1)^2+k^2}. \qquad \ldots\ldots(12)$$

The solution of these equations by simply fitting them to the experimental data for R at three frequencies is virtually impossible. The method adopted was to determine analytically (i) the frequency for minimum R, (ii) the frequency for maximum R, (iii) the magnitude of maximum R, in terms of A, ν_0 and g.

Putting the data of Yoshinaga and Oetjen (1956) into the three expressions obtained we find the equivalent values for the frequencies: $\nu_0\equiv177\cdot6\,\mathrm{cm}^{-1}$, $g\equiv5\cdot7\,\mathrm{cm}^{-1}$, and $A\equiv13\times10^4\,\mathrm{cm}^{-2}$. This represents a very narrow, but intense absorption band.

The dispersion contribution is then evaluated from equation (10), again neglecting k^2, and plotted as curve D in figure 6.

The sum of the four dispersion contributions is shown by the broken line in figure 6. It will be seen that over the range 10–20μ the line is remarkably straight, showing that the dispersion varies only as a linear function of wavelength in this region.

Figure 6 also shows that the dielectric constant will be equal to the extrapolated long wavelength value it would have if only the short wavelength bands contributed to the dispersion, at $9\cdot8\,\mu$. This is the usual definiton of the zero frequency refractive index. Hence from figure 4 we find $\epsilon=15\cdot7$ or $n_{9.8}=3\cdot96$.

In the usual optical materials the Restrahlen or free carrier absorption, if any, are sufficiently widely separated from the main absorption bands for a virtually constant value of ϵ to be reached in the region between.

The theoretical dispersion curve is plotted in figure 4, normalized at $9\cdot8\,\mu$ for the reasons given above. The overall agreement is good, and the slope at long wavelengths agrees particularly well.

It should be emphasized that the presence of the strong linear term in the dispersion means that interference fringes in a semiconductor like InSb can only be analysed by identifying the actual fringe numbers, as stated in §4. For example, if an index were calculated from the equally spaced fringes of figure 3, the erroneous value obtained would be 4·07. Methods of analysis of interference data by simply counting fringes in a given wavelength interval, and assuming specifically that there is no linear term in the dispersion, as done for example, by Rank, Bennett and Cronemeyer (1954) in their work on germanium, may therfore be suspect. Their results harmonize with those of Salzberg and Villa (1957) if a linear term is subtracted.

---

† We assume that this value may be taken as an average for the energy range concerned.

§ 6. CALCULATION OF RADIATION LIFETIME

A value of the radiation lifetime in InSb has already been given by Moss, Hawkins and Smith (1957) using the simple equation for the recombination rate which assumes no dispersion, namely

$$R=4n^2\int KQd\lambda \text{ where } Q=2\pi c\lambda^{-4}(\exp h\nu/kT-1)^{-1}$$

is the tabulated Planck function for radiation *in vacuo*. With the accurate expression where the n^2 is inside the integral and the data for n and K given here, the accurate value of R is $1{\cdot}02\times 10^{22}$.

Since the radiation lifetime is $\tau_R=n_i/2R$, where $n_i=1{\cdot}6\times 10^{16}\,\mathrm{cm}^{-3}$ at 295°K, we find $\tau_R=0{\cdot}79\,\mu$sec.

§ 7. CONCLUSIONS

Accurate values of absorption coefficient were determined from 1·5 to 7μ. They support the present theoretical estimates of the band structure.

The dispersion is found to have large contributions from both 'short' and 'long' wavelength absorption regions to such an extent that there is no region of low dispersion which can be simply identified with a specific figure for refractive index in the usual way. From the analysis of the data it is concluded that the index at 9·8μ is that which would result at infinite wavelength due to the 'short' wavelength absorption only. Its value is $n_{9.8}=3{\cdot}96$.

The fact that the calculated dispersion curve agrees with experiment establishes that the Drude theory satisfactorily explains the free carrier behaviour with an average effective mass of $0{\cdot}015m$ for intrinsic material at room temperature†, and that treating the Restrahlen band as a classical oscillator is a valid assumption.

An accurate calculation of the radiation lifetime at 295°K gives $\tau_R=0{\cdot}79\,\mu$sec.

ACKNOWLEDGMENTS

Acknowledgment is due to D. G. Avery (Royal Radar Establishment) and I. M. Ross (Services Electronics Research Laboratory) for assisting with supply of InSb, and to Miss D. Webber for helping to prepare specimens.

REFERENCES

AVERY, D. G., GOODWIN, D. W., LAWSON, W. D., and MOSS, T. S., 1954, *Proc. Phys. Soc. B*, **67**, 761.

BARDEEN, J., BLATT, F. J., and HALL, L. H., 1956, *Photoconductivity Conference* (New York : Wiley).

BRECKENRIDGE, R. G., BLUNT, R. F., HOSLER, W. R., FREDERIKSE, H. P. R., BECKER, J. H., and OSHINSKY, W., 1954, *Phys. Rev.*, **96**, 571.

---

† For n-type material where the conduction band is partially filled the effective mass will be higher (Moss 1956, Keyes *et al.* 1956).

Briggs, H. B., Cummings, R. F., Hrostowski, H. J., and Tanenbaum, M., 1954, *Phys. Rev.* **93**, 912.

Burstein, E., Picus, G. S., and Gebbie, H. A., 1956, *Phys. Rev.*, **103**, 825.

Fan, H. Y., 1956, *Rep. Progr. Phys.*, **19**, 107 (London : Physical Society).

Hrostowski, H. J., Morin, F. J., Geballe, T. H., and Wheatley, G. H., 1955, *Phys. Rev.*, **100**, 1672.

Kane, E. O., 1957, *Phys. Chem. Solids*, **1**, 249.

Keyes, R. J., Zwerdling, S., Foner, S., Kolm, H. H., and Lax B., 1956, *Phys. Rev.*, **104**, 1804.

Moss, T. S., 1952, *Photoconductivity in the Elements* (London : Butterworths Scientific Publications) ; 1956, *Photoconductivity Conference* (New York : Wiley).

Moss, T. S., Hawkins, T., and Smith, S. D., 1957, *Report of Meeting on Semiconductors* 1956 (London : Physical Society).

Oswald, F., and Schade, R. 1954, *Z. Naturforsch.*, **9a**, 611.

Rank, D. H., Bennett, H. E., and Cronemeyer, D. C., 1954, *J. Opt. Soc. Amer.*, **44**, 13.

van Roosbroeck, W., and Shockley, W., 1954, *Phys. Rev.*, **94**, 1558.

Salzberg, C. D., and Villa, J. J., 1957, *J. Opt. Soc. Amer.*, **47**, 244.

Welker, H., and Weiss, H., 1956, *Solid State Physics*, vol. 3. (New York : Academic Press).

Yoshinaga, H., and Oetjen, R. A., 1956, *Phys. Rev.*, **101**, 526.

The Overhauser Effect in Solids

By W. A. LITTLE†

Department of Natural Philosophy, University of Glasgow

MS. received 23rd October 1956, and in revised form 23rd April 1957

Abstract. A theoretical treatment is given of the Overhauser effect in solids containing two or more different nuclear species. In a rigid lattice a decrease is predicted in the macroscopic magnetic moment of one nuclear species on saturating the spin system of the other. In a lattice executing internal motion the Overhauser effect is shown to be strongly dependent upon the value of the lattice correlation time. The usual assumption of the temperature dependence of this correlation time allows one to predict the variation of the Overhauser effect with temperature. The theoretical results are compared with experiment. Good agreement is obtained for the temperature dependence of the Overhauser effect in phosphonium iodide. Agreement in ammonium fluoborate could be obtained only on the assumption that the 'internal' relaxation of the spins in the ammonium ion is small compared with the 'external' relaxation processes. At low magnetic field strengths lithium fluoride exhibited the expected behaviour of the Overhauser effect in a rigid lattice.

§ 1. Introduction

A TREATMENT of the relaxation processes in a system of two spins has been given recently by Solomon (1955). This paper represents a generalization of his treatment to a system of two types of particle with spin imbedded in a molecular environment which remains at a fixed temperature T°K. The treatment has been taken only so far as to consider the equilibrium macroscopic magnetic moment of one type of particle when the other is subjected to a strong radiation field. The resulting Overhauser effect (Overhauser 1953) is shown to depend upon the details of the spin–lattice and the spin–spin interactions. In order to carry out the generalization the treatment follows to some extent the terminology and methods used by Bloch (1956) in deriving the general Boltzmann equation of a system of nuclear spins.

§ 2. Theory

A system consisting of two types of particle, I and S, is considered. These are imbedded in a lattice which acts as a thermal reservoir at a constant temperature T°K. The particles possess both spin and magnetic moment. One wishes to compute the macroscopic magnetic moment μ_I due to the particles of spin I when the system is in thermal equilibrium in a constant magnetic field H_0 along the z-direction and also when it has reached a steady state in the presence of a radiation field which reduces μ_S to zero.

The Hamiltonian of the whole system may be written as:

$$\mathscr{H} = E + F + G \qquad \ldots\ldots(1)$$

---

† Now at the Department of Physics, University of British Columbia, Vancouver, 8, Canada.

where E represents the Zeeman energy of all the spins in the magnetic field, F represents the energy of the lattice, and G the coupling energy of the spins among themselves and with the lattice. This last term, G, is considered a small perturbation of the energies of the terms E and F. Without this perturbation the eigenstates of energy of the whole system may be described in terms of an eigenket of the form $|ms, jt, fu\rangle$. m refers to the total z-component of the spin I particles. In general this value of m will be highly degenerate. The letter s will distinguish between the many possible arrangements of the individual spins which give a particular value of m. Similarly, j will be the total z-component of the spin S particles and t the particular arrangement of the individual spins. The energy of the lattice will be given by f and the particular manner by which it is made up by u.

The representation is chosen such that the unperturbed Hamiltonian is diagonal with eigenstates specified by the quantum numbers ms, jt, fu. The state of the whole system may be described in terms of the density matrix $\langle n|\rho|n'\rangle$

$$\langle ms, jt, fu|\rho|m's', j't', f'u'\rangle.$$

Because the perturbation G is small compared with E and F the off-diagonal terms of the density matrix will be small compared with the diagonal terms. It is assumed that the approximation may be made of treating the diagonal terms as the occupation numbers $\rho(ms, jt, fu)$ of the eigenstates $|ms, jt, fu\rangle$. The rate of change of these occupation numbers may be computed from

$$\frac{d\rho_n}{dt} = \sum_{n'} \{\rho_{n'} W_{n'n} - \rho_n W_{nn'}\} \qquad \ldots\ldots(2)$$

where $W_{nn'}$ is defined as the transition probability per unit time from the state $|n\rangle$ to the state $|n'\rangle$. By using first order perturbation theory one obtains the following expression for $W_{nn'}$:

$$W_{nn'} = \frac{1}{t\hbar^2} \left| \int_0^t \langle n'|G|n\rangle \exp(-i\omega_{nn'}t')dt' \right|^2 \qquad \ldots\ldots(3)$$

(Solomon 1955).

In following the behaviour of quantities which depend only upon m and s one may use the distribution matrix σ, defined as (Bloch 1956)

$$\langle ms|\sigma|m's'\rangle = \sum_{jt} \sum_{fu} \rho \delta_{jt, j't'} \delta_{fu, f'u'}. \qquad \ldots\ldots(4)$$

The rate of change of the occupation numbers corresponding to the diagonal elements of the distribution matrix σ is obtained from equations (2) and (4)

$$\frac{d\sigma}{dt} = \sum_{fu} \sum_{jt} \left(\frac{d\rho}{dt}\right) \delta_{jt, j't'} \delta_{fu, f'u'}. \qquad \ldots\ldots(5)$$

In order to carry out the summation the perturbation G may be written as the sum of the dipole–dipole interactions between each pair of particles in the lattice.

$$G = \sum_{\tau\nu} \sum_{lk} I_l^{\tau} F_{lk}^{\tau\nu} S_k^{\nu} + \sum_{\tau\tau'} \sum_{ll'} I_l^{\tau} F_{ll'}^{\tau\tau'} I_{l'}^{\tau'} + \sum_{\tau\alpha} \sum_{lg} I_l^{\tau} F_{lg}^{\tau\alpha} J_g^{\alpha}$$
$$+ \text{terms which do not contain } I_l \qquad \ldots\ldots(6)$$

(Bloch 1956). The subscripts l, k, g refer to the lth I particle, kth S particle and gth J particle, etc.

The first term represents the coupling energy between each of the I particles and each of the S particles. The second term is the coupling energy between like spins in the group of I particles. The third term represents the coupling of the I particles directly to the lattice. The coupling to the lattice has been written as if it goes via some particles J_g which possess both spin and magnetic moment. This is convenient for the case of spin $\frac{1}{2}$ particles for this is the only way that the nuclear spin system is directly coupled to the molecular lattice.

The operator I_l^τ must obey the following matrix equation:

$$\langle ms|I_l^\tau|m's'\rangle = I_{lm}^\tau \delta_{m,m'+\tau}\delta_{s,s_{l\tau}} \qquad \ldots\ldots(7)$$

where the arrangement $s_{l\tau}$ is the same arrangement s except that the lth particle of spin I has changed its component of angular momentum along the z-direction by an amount τ.

I_{lm}^τ is the matrix element for the transition in which the z-component of I_l changes from I_m to $I_{m+\tau}$. This depends upon the initial m-state of I_l but not upon the state of the rest of the spins.

Similarly, the operator S_k^ν must satisfy the relation:

$$\langle jt|S_k^\nu|j't'\rangle = S_{kj_\nu}\delta_{j,j'+\nu}\delta_{t,t_{k\nu}} \qquad \ldots\ldots(8)$$

where the arrangement $t_{k\nu}$ is the same arrangement t' except that the kth particle of spin S has changed its z-component of angular momentum by an amount ν.

$F_{lk}^{\tau\nu}$ is an operator which acts only upon the variables of the molecular environment associated with the particles specified in the subscript.

In order to simplify the treatment both particles I and S will be assumed to have a spin of $\frac{1}{2}$. The values of τ and ν are then limited to ± 1 and 0.

The effect of the perturbation is to remove the degeneracy in both m and j. In the perturbed system each of the arrangements, s and s' will have slightly different energies. Similarly, the different arrangements t and t' will have slightly different energies. These energies will be distributed about the energy determined by m and j respectively in the unperturbed system because G is small compared with either E or F.

It shall be assumed that all the spin I particles considered lie in equivalent positions in the unit cell. Alternatively, if there are two or more non-equivalent sites in the cell, each site may be treated separately and the macroscopic magnetic moment for the whole specimen taken as the sum of the moments of each of the separate sites.

It is assumed that the occupation number $\rho(ms, jt, fu)$ may be written in the form:

$$\rho(ms, jt, fu) = \sigma(ms)\Theta(jt)\Pi(fu) \qquad \ldots\ldots(9)$$

where

$$\Pi(fu) = \frac{\exp[-E(fu)/kT]}{\sum\limits_{f'u'}\exp[-E(f'u')/kT]}. \qquad \ldots\ldots(10)$$

T is the temperature of the lattice and $E(fu)$ is the energy of the molecular state $|fu\rangle$.

Similarly,

$$\Theta(jt) = \frac{\exp[-E(jt)/kT_S]}{\sum\limits_{j't'}\exp[-E(j't')/kT_S]}. \qquad \ldots\ldots(11)$$

In this expression T_S is a parameter which describes the probability of finding the S particles in a particular state $|jt\rangle$ of energy $E(jt)$.

Similarly,

$$\sigma(ms) = \frac{\exp[-E(ms)/kT_I]}{\sum\limits_{m's'}\exp[-E(m's')/kT_I]}, \quad E(ms) \simeq m\gamma_I \hbar H_0. \qquad \ldots\ldots(12)$$

It will be assumed that the energy difference between the different arrangements s and s' is so small that $\sigma(ms)$ will be approximately equal to $\sigma(ms')$. $\sigma(m)$ will be written as the average value of $\sigma(ms)$ for all arrangements of s. Equation (5) can now be written in the form:

$$\frac{d}{dt}\sigma(m)=\sum_{fu}\Pi(fu)\sum_{jt}\Theta(jt)\left\{\sum_{m'}\sum_{j't'}\sum_{f'u'}\exp\left[\frac{-(E(j't')-E(jt))}{kT_S}\right]\right.$$
$$\left.\times\exp\left[\frac{-(E(f'u')-E(fu))}{kT}\right]\sigma(m')W_{n'n}-\sigma(m)\sum_{j't'}\sum_{f'u'}W_{nn'}\right\}. \quad \ldots\ldots(13)$$

When a steady state is reached $d\sigma(m)/dt=0$. It will be assumed that the energy differences which occur in (13) are all small compared with kT, kT_S, and kT_I so that the exponents may be expanded. The summation may now be carried out by noting the conditions imposed by equations (7) and (8). One eventually obtains

$$\frac{\gamma_I\hbar H_0}{kT_I}=\frac{\sum_{jt}\sum_{j't'}\sum_{fu}\sum_{f'u'}\left[\frac{E(jt)-E(j't')}{kT_S}+\frac{E(fu)-E(f'u')}{kT}\right](W^{+}+W^{2+})}{\sum_{jt}\sum_{j't'}\sum_{fu}\sum_{f'u'}(W^{+}+2W^{2+})}=\alpha \quad \ldots\ldots(14)$$

where the W^{+} refer to transitions in which m changes by $+1$. The W^{2+} refer to those transitions in which m changes by $+2$.

The macroscopic expectation value, Q, of any quantity $\langle ms|Q|m's'\rangle$ is obtained by taking the diagonal sum of the matrix product:

$$\langle m''s''|\sigma|ms\rangle\langle ms|Q|m's'\rangle.$$

Hence, as we have taken $\sigma(ms)$ to be approximately diagonal:

$$\mu_I=\sum_{ms}\sigma(ms)(m\gamma_I\hbar). \quad \ldots\ldots(15)$$

The degeneracy of a particular m level is the number of different arrangements s which can occur. So that

$$\sum_{s}\sigma(ms)=g_m\sigma(m) \quad \ldots\ldots(16)$$

where g_m is the degeneracy of this m level. For spin $\frac{1}{2}$ particles the degeneracy is given by the binomial coefficient $\binom{n}{\frac{1}{2}n-m}$ where the number of I particles is n.

μ_I is obtained from (15) by putting in the values of $\sigma(ms)$ from equations (12), (14) and (16).

$$\mu_I=\frac{\gamma_I\hbar\sum_{m}mg_m\,\mathrm{e}^{-\alpha m}}{\sum_{m}g_m\,\mathrm{e}^{-\alpha m}}. \quad \ldots\ldots(18)$$

The summation is carried out by making use of the binomial expansion yielding the result

$$\mu_I=-\frac{\hbar\gamma_I n}{4}\alpha. \quad \ldots\ldots(19)$$

2.1. *Thermal Equilibrium*

In the case of thermal equilibrium T_S becomes equal to the lattice temperature T°K. Furthermore, energy must be conserved within the system so that

$$E(jt)+E(fu)+E(ms)=E(j't')+E(f'u')+E(m's'). \quad \ldots\ldots(20)$$

From equation (14), α is given by

$$\alpha = \gamma_I \hbar H_0 / kT. \qquad \ldots\ldots(21)$$

From equation (19) one obtains the value of μ_I when the system is in thermal equilibrium:

$$\mu_I = -n\hbar^2 \gamma_I^2 H_0 / 4kT. \qquad \ldots\ldots(22)$$

2.2. *The Overhauser Effect*

On the application of a strong radiation field at a frequency corresponding to transitions in which j changes by ± 1, T_S rises above the temperature of the lattice. In the limit of a very strong radiation field, T_S becomes infinite. In this case equation (14) can be written:

$$\alpha = \frac{\sum\limits_{jt}\sum\limits_{j't'}\sum\limits_{fu}\sum\limits_{f'u'}\left[\frac{E(fu)-E(f'u')}{kT}\right](W^{+}+W^{2+})}{\sum\limits_{jt}\sum\limits_{j't'}\sum\limits_{fu}\sum\limits_{f'u'}(W^{+}+2W^{2+})}. \qquad \ldots\ldots(23)$$

It is now necessary to consider in detail the transition probabilities which occur in this summation. In order to calculate the transition probabilities it is necessary to know the explicit time dependence of G. The time dependence of each of the operators I, F and S may be specified by the correlation times τ_I, τ_F and τ_S respectively. These are defined in the usual way by

$$\overline{\langle I(t+\tau)\rangle\langle I^*(t)\rangle} = \overline{|\langle I(t)\rangle|^2} \exp(-\tau/\tau_I) \qquad \ldots\ldots(24)$$

$$\overline{\langle F(t+\tau)\rangle\langle F^*(t)\rangle} = \overline{|\langle F(t)\rangle|^2} \exp(-\tau/\tau_F) \qquad \ldots\ldots(25)$$

$$\overline{\langle S(t+\tau)\rangle\langle S^*(t)\rangle} = \overline{|\langle S(t)\rangle|^2} \exp(-\tau/\tau_S) \qquad \ldots\ldots(26)$$

where the bar signifies an average over t.

Two extreme conditions of the lattice corresponding firstly to a rigid lattice and secondly to a non-rigid lattice will be considered.

(*a*) *The rigid lattice.*

The rigid lattice will be defined as one in which the correlation time τ_F of equation (25) shall be long compared with either τ_I or τ_S. In this case on inserting expressions (24), (25) and (26) in (3) one obtains a negligible contribution to the time dependence of the expression from the molecular correlation time τ_F. Consequently, the transition probabilities in which f changes are small compared with those in which f remains unchanged. Hence, one obtains from (23):

$$\alpha = 0 \qquad \ldots\ldots(27)$$

and from (19)

$$\mu = 0. \qquad \ldots\ldots(28)$$

(*b*) *The non-rigid lattice.*

The non-rigid lattice will be defined as one in which the molecular correlation time τ_F shall be short compared with both the spin correlation times τ_I and τ_S. In this case the time dependent part of (3) comes from the operator F. It may then be shown that:

$$W_{nn'}{}^{+} = \sum_{lk} \frac{1}{\hbar^2} \overline{|I_l{}^{+} F_{lk}{}^{+\nu} S_k{}^{\nu}|^2} \frac{2\tau_c}{1+\omega_{ij}{}^2 \tau_c{}^2} \qquad \ldots\ldots(29)$$

where $\hbar\omega_{ij}$ is the energy difference between the states $|m-1, s'; i-\nu, t'\rangle$ and

$|ms, jt\rangle$. From (20) $\hbar\omega_{ij}$ is also the difference of energy of the states $|f, u\rangle$ and $|f', u'\rangle$.

When the explicit form of the dipole–dipole perturbation is used the operators F_{lk} may be written as:

$$\left.\begin{aligned} F_{lk}{}^{0,0} &= c_{lk}(1-3\cos^2\theta); \\ F_{lk}{}^{-1,1} = F_{lk}{}^{1,-1} &= c_{lk}(\tfrac{1}{2}-\tfrac{3}{4}\sin^2\theta); \\ F_{lk}{}^{\pm 1,0} = F_{lk}{}^{0,\pm 1} &= -c_{lk}(3/2\sin\theta\cos\theta\, e^{\pm i\phi}); \\ F_{lk}{}^{1,\pm 1} &= -\tfrac{3}{4}c_{lk}\sin^2\theta\, e^{\pm 2i\phi}, \\ \text{where } c_{lk} &= \frac{\hbar^2\gamma_I\gamma_S}{r_{lk}{}^3}. \end{aligned}\right\} \quad \ldots\ldots(30)$$

The polar angles θ and ϕ have been introduced to characterize the orientation with respect to the z-axis of the vector $\mathbf{r}_{lk}$, which joins the lth and kth particles. The fact that the transition probabilities of (29) depend through (30) upon the inverse sixth power of the separation $\mathbf{r}_{lk}$ of the interacting particles makes it possible to simplify the summation of (6) and hence of (23). In many cases a good approximation is obtained by taking the summation over nearest neighbours only.

In a polycrystalline material a further simplification is provided by the fact that the vector $\mathbf{r}$ joining any pair of particles may have any orientation with respect to the z-axis. The summation over all the particles may then be replaced by a small number of terms, where each term represents the interaction of a pair of neighbouring particles in the unit cell averaged over all orientations of the cell.

By using the same approximations as were used in obtaining (22) one can obtain μ_I. This is given by the following expression:

$$\mu_I = -n\frac{\hbar^2\gamma_I H_0}{4kT}\frac{\sum\limits_{\nu}(1+\nu\gamma_s/\gamma_I)Y_\nu}{\sum\limits_{\nu} Y_\nu} \quad \ldots\ldots(31)$$

where

$$Y_{\nu(\nu\neq 0)} = \frac{1}{\hbar^2}\sum_{lk}\overline{\langle I_l{}^{+1}F_{lk}{}^{+1,\nu}S_k{}^{\nu}\rangle^2}\,\frac{2\tau_c}{1+(\omega_I+\nu\omega_S)^2\tau_c{}^2} \quad \ldots\ldots(32)$$

$\omega_I/2\pi$ is the Larmor precession frequency of the I particles in the magnetic field and $\omega_S/2\pi$ that of the S particles.

Y_0 is given by the following expression:

$$\begin{aligned} Y_0 = \frac{2\tau_c}{\hbar^2}\Bigg[&\frac{1}{1+\omega_I{}^2\tau_c{}^2}\Big\{\sum_{lk}\overline{\langle I_l{}^{+1}F_{lk}{}^{+1,0}S_k{}^0\rangle^2} + \sum_{ll'}\overline{\langle I_l{}^{+1}F_{ll'}{}^{+1,0}I_{l'}{}^0\rangle^2}\Big\} \\ &+ \frac{2}{1+4\omega_I{}^2\tau_c{}^2}\sum_{ll'}\overline{\langle I_l{}^{+1}F_{ll'}{}^{+1,+1}I_{l'}{}^{+1}\rangle^2} \\ &+ \sum_{\alpha=-1}^{\alpha=+1}\frac{1}{1+(\omega_I+\alpha\omega_\alpha)^2\tau_c{}^2}\sum\overline{\langle I_l{}^{+1}F_{lg}{}^{+1,\alpha}J_g{}^{\alpha}\rangle^2}\Bigg]. \end{aligned} \quad \ldots\ldots(33)$$

The summations of (32) and (33) are simplified by noting that

$$\left.\begin{aligned} \overline{|J^{\pm 1}|^2} &= \tfrac{2}{3}J(J+1)(2J+1) \\ \overline{|J^0|^2} &= \tfrac{1}{3}J(J+1)(2J+1) \\ \overline{|F_{lk}{}^{+1,+1}|^2} = \overline{|F_{lk}{}^{+1,0}|^2} &= \tfrac{3}{10}C_{lk}{}^2 \\ \overline{|F_{lk}{}^{+1,-1}|^2} &= \tfrac{1}{20}C_{lk}{}^2. \end{aligned}\right\} \quad \ldots\ldots(34)$$

In (6) the interaction of the particles I with the lattice was written as if the interaction went via some other particle J_g possessing a magnetic moment. $\omega_\alpha/2\pi$ is the Larmor precession frequency of the J particles in the magnetic field.

It is convenient at this point to consider the validity of the result contained in equation (31). It has been assumed that:

(*a*) The molecular lattice remains in thermal equilibrium at a temperature T°K.

(*b*) The motion of the molecular lattice is random and may be described in terms of a single correlation time τ_c.

(*c*) This correlation time is short compared with the spin correlation time.

(*d*) The mean width of a particular m or j energy level is small compared with the respective separation of these levels in the magnetic field.

(*e*) The separation of these levels in the magnetic field is small compared with kT.

(*f*) The particles I and S have spin $\frac{1}{2}$.

(*g*) The material is in powder or polycrystalline form.

(*h*) The spin I and spin S systems can each be described by equations (11) and (12).

In the case where the particles J_g are nuclei with spin $\frac{1}{2}$ the assumption (*a*) may not be justified. The reason for this is that the Overhauser effect may upset the thermal equilibrium distribution of the J particles. If J has a spin greater than $\frac{1}{2}$ the interaction with the lattice via the quadrupole moment of the nucleus usually will be stronger than the magnetic spin–spin interaction between the particles. In this case the J particles remain in thermal equilibrium with the lattice and the assumption (*a*) is justified.

§ 3. EXPERIMENTAL INVESTIGATION

In order to compare the results of § 2 with experiment a nuclear magnetic resonance spectrometer was used. A permanent magnet provided a uniform magnetic field of 3600 oersteds between polepieces 4 in. in diameter and 1⅜ in. apart. The specimen was contained in a 3 cm^3 glass specimen tube which fitted into the tank coil of an oscillator-detector of the type described by Pound and Knight (1950). This coil replaced the receiver coil in a crossed coil bridge box designed on the lines of that described by Weaver (1953). A variable frequency oscillator could deliver radio-frequency power to the transmitter coil in the bridge. The transmitter coil could provide a rotating magnetic field of up to 1/3 oersted at a frequency suitable for saturating the spin levels of the S particles. The bridge was adjusted to balance out any signal generated in the detector coil by the transmitter coil. The magnetic field at the specimen was modulated at 300 c/s by two pancake coils cemented on either side of the bridge box. A phase-sensitive detector displayed the derivative with respect to the magnetic field of the nuclear magnetic absorption on a pen-recorder. In this method of display the recorded signal is proportional to the macroscopic magnetic moment of the observed nuclear specimen, provided that the line width remains constant.

A cryostat operating on the heat leak principle allowed the temperature of the box to be varied from −90°C to +90°C. This provided a convenient means of varying the molecular correlation time of the lattice. The temperature of the specimen was measured by a thermocouple and galvanometer.

At each temperature a trace of the absorption signal was taken. The variable frequency oscillator was then tuned to the Larmor precession frequency of the S particles. The oscillator power was adjusted to ensure that the spin levels were 'saturated'. A second trace was then taken of the absorption. The ratio of these two signals was plotted against the temperature.

§ 4. Results

(*a*) *The rigid lattice.*

Lithium fluoride and triphenyl-phosphine were chosen as two materials which might be expected to satisfy the conditions of a rigid lattice. An unsuccessful attempt was made to detect the reduction in the size of the macroscopic magnetic moment as predicted by equation (28). The spin–lattice relaxation time of the protons in triphenyl-phosphine was found to be approximately 20 seconds at room temperature. The failure of the experiment was attributed to the presence of a small amount of paramagnetic impurity in both compounds. These impurities serve to maintain the nuclear spins in thermal equilibrium. The nuclear spins in the immediate neighbourhood of the impurity are rapidly relaxed by the spin–spin interaction. Energy diffuses from these centres to the rest of the nuclear spins (Bloembergen 1949). Energy is transferred by the mutual flip of a pair of nuclear spins. The rate at which this energy diffuses from the impurity centres will be practically independent of the applied field. However, the transition probabilities which are required for (28) are given by an expression of the form (29) with the spin correlation times replacing the molecular correlation time. These probabilities will be large for small ω_{ij}. Thus, at low magnetic fields the predictions of (28) should not be masked by the presence of impurities.

Low field experiment. A single crystal of lithium fluoride was found to have a spin–lattice relaxation time of 45 seconds for the fluorine nuclei in a field of 3600 oersteds. The crystal was allowed to reach thermal equilibrium in the magnetic field. It was then removed to a position in the fringing field of the magnet where the field was 70 oersteds. After a measured interval of time it

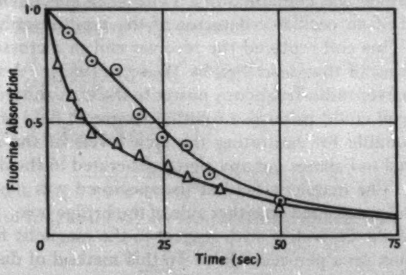


Figure 1. Circles : Decay of the fluorine macroscopic magnetic moment in a single crystal of lithium fluoride with time after reducing the magnetic field from 3600 oersted to 70 oersted. Triangles : The same effect but with the lithium spin system initially completely saturated.

was then returned to the spectrometer and the absorption noted. This was then repeated with the ^{7}Li resonance saturated. Figure 1 illustrates the result obtained. A reduction of the fluorine macroscopic magnetic moment occurs when the ^{7}Li spin system is saturated. This is in agreement with the prediction of equation (28). At 400 oersteds the effect is almost completely masked by the presence of paramagnetic impurities.

(b) The non-rigid lattice.

Phosphonium iodide and ammonium fluoborate were chosen as two materials which might be expected to satisfy the conditions of a non-rigid lattice. It is well known that the ammonium ion in an ammonium salt can execute hindered rotation at temperatures well below the melting point of the solid. One might expect a similar behaviour of the phosphonium ion. The behaviour of the Overhauser effect as predicted by equation (31) depends strongly upon the molecular correlation time τ_c. In the case of hindered rotation the temperature dependence of the correlation time is usually found to have the form:

$$\tau_c = \tau_0 \exp(V/T) \qquad \ldots\ldots(35)$$

where T is the absolute temperature and τ_0 and V are constants. Inserting this expression in (31), the temperature dependence of the Overhauser effect may be predicted. This has been done for the chemicals mentioned above for the nuclei ^{31}P, ^{1}H and ^{19}F, all of which have a spin $\frac{1}{2}$.

Phosphonium iodide. The ^{31}P resonance was observed in solid phosphonium iodide over a temperature range from $-70°$c to $+34°$c. The ^{1}H resonance was subjected to a strong radiation field. In the phosphonium ion the ^{31}P nucleus lies centrally in the PH_4 group. The P–H bond length is 1·42 Å and the P–I separation is 3·90 Å (Wells 1950). The Overhauser effect was calculated using the P–H and P–I interactions. It has been assumed that the molecular correlation time has the following temperature dependence:

$$\tau_c = 9{\cdot}8 \times 10^{-16} \exp(4200/T).$$

The experimental points and the theoretical curve are compared in figure 2. A transition occurs in the width of the ^{31}P resonance line at $-75°$c. Above this temperature the line is about 3·4 gauss wide.

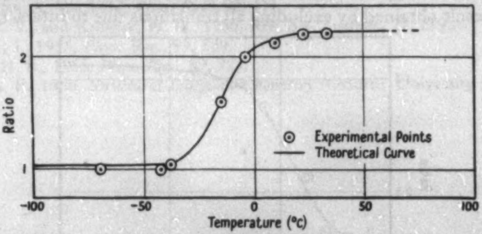


Figure 2. Phosphonium iodide. Phosphorus resonance. Variation of the Overhauser effect with temperature.

Ammonium fluoborate. NH_4BF_4 has the crystal structure of $BaSO_4$ (Wells 1950). The hydrogen atoms are bound tetrahedrally to the nitrogen, and the fluorine atoms are bound similarly to the boron. The Overhauser effect has been observed on the ^{1}H resonance with the ^{19}F resonance 'saturated'. The effect has been calculated taking into account the N–H, H–H and the H–F interactions. The N–H bond length has been taken as 1·04 Å. Twelve fluorine nuclei lie at approximately 3·06 Å from the nitrogen nucleus (Bragg 1937). The orientation of the ammonium ion is not known. The average interaction of all possible orientations of the ammonium ion was calculated from the dimensions of the unit cell and the crystal structure. The temperature dependence of the molecular correlation time was taken as:

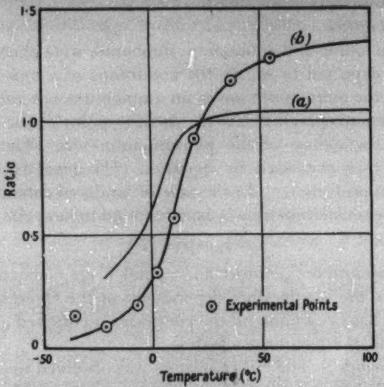


Figure 3. The variation of the Overhauser effect with temperature in ammonium fluoborate. The proton resonance is observed with and without the fluorine spin system being saturated. (*a*) Theoretical curve computed using all interactions. (*b*) Theoretical curve computed excluding relaxation processes due to identical particles in the ammonium ion.

$$\tau_c = 2{\cdot}6 \times 10^{-15} \exp(4600/T).$$

The results obtained are given in figure 3. Curve 3 (*a*) gives the theoretically calculated effect taking into account all the interactions mentioned above. Curve 3(*b*) is the result obtained by excluding all transitions due to interactions with the

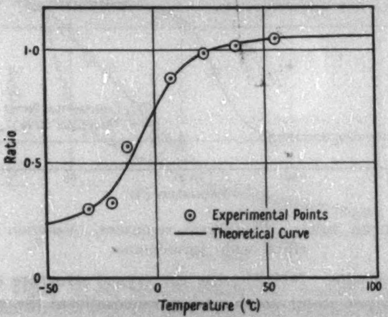


Figure 4. Variation of the Overhauser effect in ammonium fluoborate with temperature. Observation of the fluorine spin polarization with the proton spin system saturated.

other protons within the ammonium ion. It is not obvious why these transition probabilities should be so small that they may be neglected compared with the transitions due to other neighbouring nuclei. In order to understand this point a detailed treatment of the relaxation processes of a hindered rotator would be required.

Figure 4 shows the experimental and theoretical results obtained with the ^{19}F resonance with the ^{1}H resonance saturated. The B–F bond length was taken as 1·40 Å. Again the F–F interactions in the BF_4 ion have been excluded. The molecular correlation time obtained from the ^{1}H results was used to compute this result.

§ 5. CONCLUSION

It is concluded that the result of (31) with (35) gives an explanation of the general trend of the temperature dependence of the Overhauser effect in crystalline solids. This result could be a valuable addition to the study of line widths for obtaining information on crystal structure and internal motion. In the rigid lattice the theory predicts a depolarization of the nuclear spins I upon saturation of the spin S system. This has been verified at low fields where paramagnetic impurities do not mask the effect.

ACKNOWLEDGMENTS

I wish to thank Messrs. Albright and Wilson for kindly supplying a sample of phosphonium iodide; and Barr and Stroud Ltd. for supplying the single crystal of lithium fluoride. To Dr. G. A. P. Wyllie and the members of the Solid State Group I am indebted for many useful discussions. I wish to thank the Shell Petroleum Company for a research scholarship which was held for the early period of this research.

REFERENCES

BLOCH, F., 1956, *Phys. Rev.*, **102**, 104.
BLOEMBERGEN, N., 1949, *Physica*, **15**, 386.
BRAGG, W. L., 1937, *Atomic Structure of Minerals*, (Oxford : University Press).
OVERHAUSER, A. W., 1953, *Phys. Rev.*, **92**, 411.
POUND, R. V., and KNIGHT, W. D., 1950, *Rev. Sci. Instrum.*, **21**, 219.
SOLOMON, I., 1955, *Phys. Rev.*, **99**, 559.
WEAVER, H. E., 1953, *Phys. Rev.*, **89**, 923.
WELLS, A. F., 1950, *Structural Inorganic Chemistry* (Oxford: University Press).

Hyperfine Structure Coupling Constant of the 6s6p $^3P_1$ Level of ^{199}Hg by Double Resonance

By G. S. BOGLE†, J. N. DODD and W. L. McLEAN
Department of Physics, University of Otago, New Zealand

Communicated by R. R. Nimmo ; MS. received 31*st December* 1956, *and in revised form* 2*nd April* 1957

Abstract. A double resonance experiment on natural mercury is described. The value of the hyperfine structure coupling constant of the 6s6p$^3P_1$ level of ^{199}Hg obtained is $0{\cdot}497 \pm 0{\cdot}014$ cm^{-1}.

§ 1. Introduction

ALTHOUGH a number of optical determinations of the hyperfine structure of the $^3P_1$ level of the odd isotopes of mercury have been made by conventional high resolution interferometry (Schüler and Schmidt 1935, Murakawa and Suwa 1950, Burns and Adams 1952, Murakawa 1955) the accuracy of the results obtained was limited by the widths of the spectral lines, interferometer line widths and, where natural mercury was used, by the overlapping of the line components of different isotopes. In a recent investigation by the magnetic scanning method of Bitter *et al.* (1954) the accuracy was limited by the Doppler broadening in both the source lamp and the absorbing vapour. The work of Brossel and Bitter (1952) on double resonance in mercury has shown that it is possible to attain a double resonance line width hardly greater than the natural width of the $^3P_1$ level, namely 2 Mc/s. In using this technique to determine the hyperfine structure coupling constants of ^{199}Hg and ^{201}Hg one must observe transitions either between the zero field hyperfine sub-levels or between the states of the atom in a magnetic field of sufficient intensity to destroy the strong coupling between I and J. The former method requires a variable frequency oscillator of about 20 000 Mc/s, delivering a few tens of watts, while in the latter method an oscillator of fixed frequency can be used, and the resonance found by varying the magnetic field.

In the experiment described here a microwave frequency of 4250 Mc/s was used. This required magnetic fields of about 2–3 kilogauss which were sufficient to produce an incomplete Back–Goudsmit effect.

§ 2. Apparatus

About 100 watts of microwave power generated by a CV80 klystron at 7 cm wavelength was transmitted by S-band waveguide and coaxial line to a rectangular cavity resonating in the H_{210} mode as shown in figure 1. A determination of the frequency by means of a standing-wave meter gave the result $\nu = 4250 \pm 5$ Mc/s. The fused quartz resonance cell, evacuated and filled with mercury vapour, was placed in the resonant cavity as shown in figure 2; the tail of the resonance cell being cooled with ice-water. In figure 2 the direction of the steady field is perpendicular to the plane of the drawing.

---

† Now at Division of Physics, National Standards Laboratory, Sydney, Australia.

The exciting optical radiation was reflected into the cavity through a hole in its side so that it passed through the vapour along the steady field direction, thus producing σ irradiation. The hole, which was about 1 cm in diameter, was crossed by several parallel strips of copper foil set edge on to the incoming light and parallel to the current flow in the cavity wall in order to prevent radiation of the microwave power.

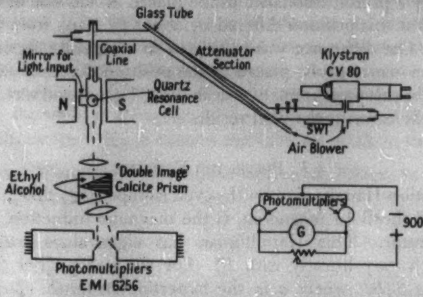

Glass Tube
Mirror for Light Input
Coaxial Line
Attenuator Section
Klystron CV 80
N
S
Quartz Resonance Cell
SWT
Air Blower
Ethyl Alcohol
'Double Image' Calcite Prism
Photomultipliers
G
900 v
Photomultipliers EMI 6256

Figure 1. Experimental arrangement.

The light source was the inner tube of a high pressure mercury lamp excited by a 170 Mc/s oscillator of 40 watt output. This lamp was found to give a double resonance signal/noise ratio ten times as large as that of a forced-air-cooled standard high pressure lamp operating from the 50 c/s mains.

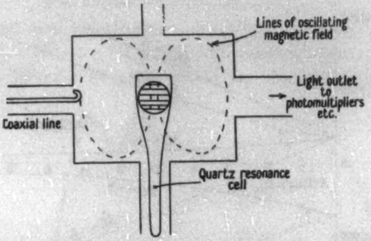

Lines of oscillating magnetic field
Light outlet to photomultipliers etc.
Coaxial line
Quartz resonance cell

Figure 2. Microwave cavity and resonance cell. Figures 1 and 2 are sections in perpendicular planes.

The changes in the intensities of the π and σ components of the re-emitted resonance radiation were detected by the photomultipliers and circuit shown in figure 1. The out of balance current (proportional to $\Delta I_\pi - \Delta I_\sigma$) was registered by a galvanometer of 15 seconds period. The photomultiplier currents were each a few microamperes.

The steady magnetic field was produced by an electromagnet of pole diameter 10 cm. A gap as large as 7 cm between the pole pieces had to be used in order to accommodate the deflecting mirror as well as the cavity. Shimming rings were used to reduce the inhomogeneity to less than 0·5 gauss over 1 cm^3 at about 3 kilogauss. Sufficient homogeneity and sensitivity were achieved to observe the double-peaked resonance curve of the even isotopes discussed by Brossel and Bitter (1952). From the separation of the peaks it was deduced that the amplitude of the oscillating magnetic field in the microwave cavity was between 2 and 4 gauss.

The actual experimental results were obtained by observing the steady magnetic field measured by a proton resonance probe placed at one side of the gap. The value of the field at this position differed by about 30 gauss from the field in the resonance cell. The difference was measured in a separate experiment using a second proton resonance probe placed in the position normally occupied by the cell. The error in this measurement was about ± 1 gauss and was responsible for a major part of the error in the final result.

§ 3. PREDICTED EFFECTS

The perturbation Hamiltonian for the even isotopes is $g_J\,\beta B J_z$, where g_J is the Landé g factor, β the Bohr magneton, B the magnetic induction, J_z the angular momentum operator. This Hamiltonian has eigenvalues corresponding to $M_J = \pm 1, 0$ which vary linearly with B. For the isotope 199 the operator is $a\mathbf{I}\,.\,\mathbf{J} + g_J\beta B J_z - g_I\beta_N B I_z$, where a is the hyperfine structure coupling constant and g_I, β_N, I_z refer to the nucleus. The eigenvalues of this operator plotted against B are given in figure 3 using the values $a = 14\,790$ Mc/s (Brix and Kopfermann 1952) and $g_J = 1{\cdot}4838$ (Brossel and Bitter 1952). The perturbation operator for the isotope 201 is similar to that for 199 but has an additional term describing the electric quadrupole interaction. Approximate numerical solutions for the energy eigenvalues of the $^3P_1$ level of ^{201}Hg have been obtained using data given by Brix and Kopfermann. Up to the present we have not observed any double resonance signals which can be attributed to this isotope and so detailed calculations have not yet been undertaken. However, a rough estimate indicates that all the possible transitions occur above 2950 ± 30 gauss. The selection rule for magnetic dipole transitions is $\Delta M = \pm 1$, where $M = M_I + M_J$ for odd isotopes, and $M = M_J$ for even isotopes. The two allowed transitions in the even isotopes occur at the same field strength given by $h\nu = g_J\beta B_0$. For ^{199}Hg there are four allowed transi-

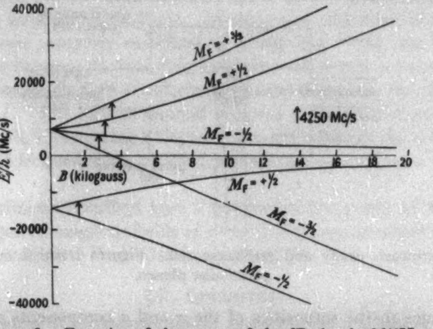


Figure 3. Energies of the states of the $^3P_1$ level of ^{199}Hg.

tions occurring at different values of the field. These are shown in figure 3 for a frequency of 4250 Mc/s.

The measured double resonance effect depends on the changes in the intensities of the π and σ components of the re-emitted radiation which occur when the magnetic field passes through resonance. These changes depend on the difference in population of the two states concerned, the microwave transition probability, the optical transition probabilities, the direction of viewing and the isotopic abundances. The difference in population of the states depends only on the probabilities for the optical transitions leading to the states since the populations of the ground states are assumed to be equal.

The double resonance effect per atom for the four allowed transitions in ^{199}Hg is given in the table as a fraction of the effect per atom of the even isotopes. The calculated effects for ^{199}Hg in natural mercury are given in the sixth column.

Predicted Magnitudes of Double Resonance Effects in ^{199}Hg

(1)	(2)	(3)	(4)	(5)	(6)
Even isotopes	2041	100	100	100	100
$F=\frac{3}{2}$ $M_F=\frac{3}{2}\rightarrow\frac{1}{2}$	3473	7·4	0	3·1	0·8
$\frac{1}{2}\rightarrow-\frac{1}{2}$	3006	0	25	1·3	0·3
$-\frac{1}{2}\rightarrow-\frac{3}{2}$	2835	7·4	25	11	3
$F=\frac{1}{2}$ $M_F=\frac{1}{2}\rightarrow-\frac{1}{2}$	1535	0	25	0·3	0·08

(1) Participating states of excited level; (2) resonance field (gauss) $\nu=4250$ Mc/s; (3)–(5) double resonance effect per atom: (3) zero field, (4) high fields, (5) at resonance field for $\nu=4250$ Mc/s; (6) total effect in natural mercury.

The effects for ^{201}Hg have not been calculated fully but are expected to be about one third of those for ^{199}Hg by reason of the smaller abundance and the larger number of ground states.

§ 4. Experimental Results

We have observed two resonances: a strong resonance due to the even isotopes at a field $B_e=2041\cdot6\pm1\cdot4$ gauss, and another resonance, about 5% as strong, at a field $B=2836\cdot3\pm1\cdot6$ gauss which we have identified without doubt

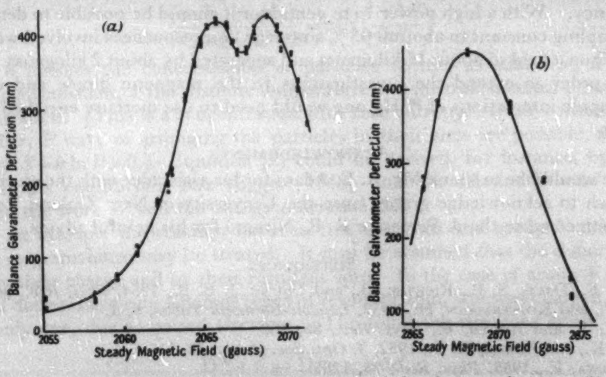


Figure 4. Experimental double resonance curves: (*a*) for the even isotopes, (*b*) for the isotope 199.

as the $F=3/2$, $M_F=-3/2 \rightarrow F=3/2$, $M_F=-1/2$ transition in ^{199}Hg. (All resonances from ^{201}Hg are expected to occur above 2950 ± 30 gauss.) Figure 4 shows typical experimental curves for the two resonances ; the deflections for the ^{199}Hg resonance were magnified by increasing the gain of the photomultipliers. The points on the graph show readings taken at 15 second intervals at fixed values of the steady field. The centre of the ^{199}Hg resonance was determined by a number of independent experiments.

The resonance condition for the observed transition in ^{199}Hg gives

$$a=\frac{h\nu(B-B_e)}{\frac{3}{2}B_e-B}$$

if one neglects the direct interaction between the steady field and the nuclear moment, but in the present experiments the error involved is negligible. It should be noted that if the experiments were performed at lower frequencies and fields, the value of B would tend to $\frac{3}{2}B_e$ and thus make the determination of a very inaccurate.

Using the values quoted above we obtain for the hyperfine coupling constant $a=0{\cdot}497 \pm 0{\cdot}013$ cm^{-1}. The major part of the error is due to the difficulty in obtaining an accurate result for the difference $\frac{3}{2}B_e-B$.

We give below results for a obtained by other authors or calculated from their results.

Burns and Adams 1952 (optical)	0·491 cm^{-1}
Bitter *et al.* 1954 (magnetic scanning)	0·479 ± 0·014 cm^{-1}
Brix and Kopfermann 1952 (review, optical)	0·4929 cm^{-1}
Present work	0·497 ± 0·013 cm^{-1}

§ 5. Conclusion

In the present experiments the use of the double resonance technique for the determination of the hyperfine structure coupling constant was limited by the inaccuracy of the magnetic field measurements. Greater accuracy in the value of the coupling constant could be obtained in two ways : by using a larger magnet producing a more homogeneous field or by using a microwave system of higher frequency. With a high power 1 cm generator it should be possible to determine the coupling constant to about 0·05% since the two resonances involved would be centred on a field of about 15 kilogauss and separated by about 2 kilogauss.

In order to extend the investigations to the magnetic dipole and electric quadrupole interactions of ^{201}Hg one would need to use mercury enriched in this isotope.

Acknowledgments

We would like to thank Mr. A. E. Adam for his assistance with the electronics. We wish to acknowledge grants from the University of New Zealand Research Committee and to thank Professor R. R. Nimmo for his helpful advice.

References

Bitter, F., Davis, S. P., Richter, B., and Young, J. E. R., 1954, *Phys. Rev.*, **96**, 1531.
Brix, P., and Kopfermann, H., 1952, *Landolt-Börnstein Tables*, **5**, 3.
Brossel, J., and Bitter, F., *Phys. Rev.*, **86**, 308.
Burns, K., and Adams, K. B., 1952, *J. Opt. Soc. Amer.*, **42**, 716.
Murakawa, K., 1955, *Phys. Rev.*, **98**, 1285.
Murakawa, K., and Suwa, S., 1950, *J. Phys. Soc. Japan*, **5**, 429.
Schüler, H., and Schmidt, T., 1935, *Z. Phys.*, **98**, 239.

RESEARCH NOTES

The Mass Action Laws for the Reactions between Free Carriers and Impurities in Semiconductors considering the Electron Spin

By F. W. G. ROSE
Research Laboratory, British Thomson-Houston Co. Ltd., Rugby

MS. received 26*th April* 1957

In addition to other treatments (e.g. Landsberg 1952, Guggenheim 1953), Crawford and Holmes (1954) have derived the effect which the electron spin has on the recombination of electrons with impurities by treating free electrons, charged and uncharged impurities as chemical entities. The chemical aspect may be further developed by considering particles with different electron spin as separate entities, to allow the application of a general rule which does not involve a knowledge of the partition function. This rule enables us also to consider certain properties of the degenerate case by treating unoccupied levels in the conduction band and occupied levels in the valence band as reacting particles. (The concentrations of these particles are usually small compared with the density of lattice points so that the laws of dilute solutions will apply.) However, it must be emphasized that the final results depend on the values of the 'effective density' of energy levels in the conduction or valence bands respectively, which are not well defined in the degenerate case. In the non-degenerate case these densities are constants and allow a useful interpretation of the results.

According to the rule, the mass action law for a reaction

$$\mathrm{A}+\mathrm{B} \rightleftharpoons \mathrm{C}+\mathrm{D}+\Delta\mathscr{E} \qquad \text{......(1)}$$

between the particles A, B, C, D, where $\Delta\mathscr{E}$ is the energy difference between the right and the left side of the equation, can be written in the following form

$$\frac{(n_{\mathrm{C}}/\tilde{n}_{\mathrm{C}})(n_{\mathrm{D}}/\tilde{n}_{\mathrm{D}})}{(n_{\mathrm{A}}/n_{\mathrm{A}})(n_{\mathrm{B}}/\tilde{n}_{\mathrm{B}})}=\exp\left(\frac{-\Delta\mathscr{E}}{\boldsymbol{k}T}\right) \qquad \text{......(2)}$$

where n denotes the concentration of particles written as subscript, $\boldsymbol{k}$ is the Boltzmann constant, T the absolute temperature and the tilde denotes a maximum concentration. (This is a concentration with zero entropy. If the concentration is smaller, P ways of arranging the particles in their sites are possible, and the entropy $S=\boldsymbol{k}\ln P\neq 0$.) Equation (2) could be derived, for instance, by using van't Hoff's equilibrium box (Eggert 1948) and the relation $W=\Delta\mathscr{E}-T\Delta S$ between the work W and the energy difference $\Delta\mathscr{E}$.

As an example the reaction between free electrons and group V donors in silicon or germanium may be treated. It may be assumed that the donors differ only by their charge and by their resultant spin. In the case of group V donors neutral donors have two different types of resultant spin, whereas charged donors have only one, due to paired electrons. The following reactions apply

$$\overrightarrow{\mathrm{D}}^{\times}+\overrightarrow{\mathrm{v}} \rightleftharpoons \mathrm{D}^{+}+\overrightarrow{\mathrm{e}}+\Delta\mathscr{E} \qquad \text{......(3)}$$

$$\overleftarrow{\mathrm{D}}^{\times}+\overleftarrow{\mathrm{v}} \rightleftharpoons \mathrm{D}^{+}+\overleftarrow{\mathrm{e}}+\Delta\mathscr{E} \qquad \text{......(4)}$$

where D$^\times$, D$^+$, e, v are neutral donors, charged donors, free electrons and unoccupied energy levels in the conduction band respectively and where the arrow denotes the spin. From equations (1) and (2) the mass action law for reaction (3) is

$$\frac{(n_{D^+}/N_D)(\vec{n}_e/\vec{N}_c)}{(\vec{n}_{D^\times}/N_D)(\vec{n}_v/\vec{N}_c)} = \frac{\vec{n}_e n_{D^+}}{(\vec{N}_c - \vec{n}_e)\vec{n}_{D^\times}} = \exp\left(\frac{-\Delta\mathscr{E}}{kT}\right) \qquad \ldots\ldots(5)$$

where N_D is the total concentration of donors and where $\vec{N}_c$ is an 'effective' density of energy levels in the conduction band with $\rightarrow$ spin. It is assumed that $\vec{N}_c = \overleftarrow{N}_c = N_c/2$. From Fermi statistics N_c is defined by

$$n_e = N_c\{1 + \exp[(\mathscr{E}_c - \mathscr{E}_F)/kT]\}^{-1},$$

where $\mathscr{E}_F$ is the Fermi level and $\mathscr{E}_c$ an 'effective level' in the conduction band. In the non-degenerate case N_c becomes constant for a given temperature and $\mathscr{E}_c$ is effectively the lowest level in the conduction band. A similar mass action law applies to equation (4). Assuming equal probabilities for both types of spin in the donors

$$\frac{n_e n_{D^+}}{(N_c - n_e)n_{D^\times}} = \frac{1}{2}\exp\left[\frac{-(\mathscr{E}_c - \mathscr{E}_D)}{kT}\right] \qquad \ldots\ldots(6)$$

is obtained, where $\mathscr{E}_D$ is the energy level of the donors. Equation (6) is in agreement with the references given at the end.

The assumption of equal probabilities for both types of spin in the donors may not be valid if the recombination cross sections depend on the spin orientation. However, this assumption is not necessary in the non-degenerate case ($N_c - n_e \simeq N_c$), so that for this case no effect of different recombination cross sections can be expected.

The other mass action laws applying to group V donors and group III acceptors in silicon and germanium are (assuming again that donors or acceptors differ only by their charge and their spin, and have equal probabilities for both types of spin):

$$\frac{n_h n_{D^\times}}{(N_v - n_h)n_{D^+}} = 2\exp\left[\frac{-(\mathscr{E}_D - \mathscr{E}_v)}{kT}\right] \qquad \ldots\ldots(7)$$

$$\frac{n_h n_{A^-}}{(N_v - n_h)n_{A^\times}} = \frac{1}{2}\exp\left[\frac{-(\mathscr{E}_A - \mathscr{E}_v)}{kT}\right] \qquad \ldots\ldots(8)$$

$$\frac{n_e n_{A^\times}}{(N_c - n_e)n_{A^-}} = 2\exp\left[\frac{-(\mathscr{E}_c - \mathscr{E}_A)}{kT}\right] \qquad \ldots\ldots(9)$$

whence

$$\frac{n_e n_h}{(N_c - n_e)(N_v - n_h)} = \exp\left[\frac{-(\mathscr{E}_c - \mathscr{E}_v)}{kT}\right] \qquad \ldots\ldots(10)$$

where $\mathscr{E}_A$ is the energy level of the acceptors, $\mathscr{E}_v$ is an 'effective energy level' (highest level in the non-degenerate case) in the valence band and where h, A$^\times$, A$^-$ denote free holes, neutral acceptors, charged acceptors respectively. N_v is the 'effective' density of energy levels in the valence band. In the non-degenerate case $n_h \ll N_v$ and

$$N_c = 2{\cdot}5 \times 10^{19}(m_e/m)^{3/2}(T/300)\ \text{cm}^{-3}$$

$$N_v = 2{\cdot}5 \times 10^{19}(m_h/m)^{3/2}(T/300)\ \text{cm}^{-3}$$

where m_e, m_h are the effective mass of electrons in the conduction band (free electrons) and of holes in the valence band (free holes) respectively, and m is the electron rest mass.

ACKNOWLEDGMENTS

I should like to express my thanks to my colleagues, Mr. E. L. Heasell and Dr. P. A. Lee in the Research Laboratory of the British Thomson-Houston Co., Rugby, to Dr. A. Hoffmann, Siemens-Schuckertwerke, Pretzfeld, for helpful discussions and to Mr. I. Williams of the British Thomson-Houston Co. for his encouragement and advice.

REFERENCES

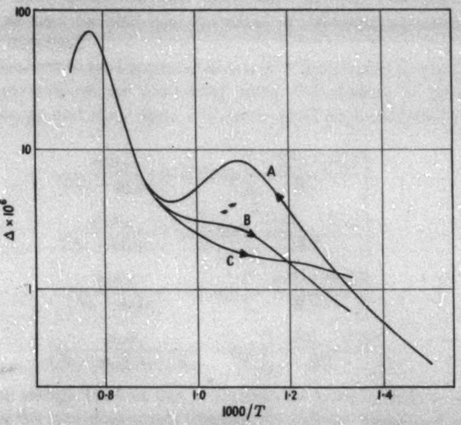
CRAWFORD, J. H., Jr., and HOLMES, D. K., 1954, *Proc. Phys. Soc.* A, 67, 294.
EGGERT, J., 1948, *Lehrbuch der Physikalischen Chemie* (Leipzig : Hirzel), p. 402.
GUGGENHEIM, E. A., 1953, *Proc. Phys. Soc.* A, 66, 121.
LANDSBERG, P. T., 1952, *Proc. Phys. Soc.* A, 66, 662.

LETTERS TO THE EDITOR

Internal Friction of Single Crystal Silicon from 25°c to 1100°c

The internal friction of silicon is being investigated at 100 kc/s in the temperature range 25°c to 1100°c. This letter reports preliminary results on some crystals of about 8Ω cm resistivity. The measuring apparatus employs electrostatic drive and frequency modulation detection on the principle described by Bordoni (1947). Decay curves are recorded and the logarithmic decrement Δ calculated from them. Strain amplitudes of measurement are less than 10^{-6}, and in general little amplitude dependence of Δ is found in this range. The silicon crystals are half-wavelength bars, 5 mm × 5 mm × 42 mm, with the long axis in a ⟨111⟩ direction. They are nodally supported in high vacuum by 0·05 mm tungsten wires.

The temperature dependence of Δ is shown in the figure where ln Δ is plotted against inverse absolute temperature. Two maxima are seen, at 650° and 1020°c. Curve A shows the ' normal ' behaviour, which is usually exhibited when measurements are made at a continuously increasing temperature. The shape of this peak varies to a limited extent between crystals, and with ageing, but its position remains constant. A quenching effect is observed if measurements are made as the specimen cools, and the height of the peak can be reduced in this way. Curves B and C show the behaviour after cooling from 1000° c over periods of 2 hours and ½ hour respectively. Subsequent recovery appears in general to occur



Decrement of 8 Ω cm silicon crystals. Behaviour, A during heating ; B and C, during cooling.

rapidly, and curve A can be retraced on immediate reheating. Internal friction at room temperature is very low indeed, the Δ of one crystal being 9×10^{-8}, corresponding to a mechanical Q of $3{\cdot}5 \times 10^{7}$. Experimental points are too numerous to plot individually on the graph. Their mean deviation from the curves drawn is about 2%.

Pre-straining has been carried out before one set of measurements, the crystal being twisted through 13° in nitrogen–hydrogen at 1000°c. The primary effect of this treatment is to raise the height of the 1020°c peak and the width of the 650°c peak, but after 48 hours anneal at 1100°c both peaks return to near their original form. There is, however, a more permanent effect of pre-straining, that the quenching effect is inhibited, curve A being produced on cooling as well as heating.

The smooth temperature dependence of Δ sometimes has sharp peaks superimposed. These are very unrepeatable and are associated with a strong amplitude dependence of Δ. Although it is possible that they are due to slip processes in the crystal, it is more likely they represent sporadic losses at the specimen support, and so have been ignored in the figure.

The absence of amplitude dependence in the normal behaviour of the crystals indicates that the internal friction is due to a relaxation effect. Analysis of curve A shows that it can be fitted quite accurately with two temperature activated relaxation peaks having single activation energies, plus a very low constant background. Activation energies calculated in this way are 52 kcal mole^{-1} for the 1020°c peak and 23 to 30 kcal mole^{-1} for the 650°c peak. This method of estimation is not so basically sound as the frequency-shift method, but is regarded as quite valuable since the fit to the theoretical curve is good.

More data are being collected on vacuum straining and crystals of differing impurity content, and it would be premature to attempt detailed discussion and interpretation of the results here. There are, however, several points which indicate that the upper and lower peaks are produced by impurity and dislocation processes respectively.

The activation energy of the lower peak is large compared with that normally associated with the Bordoni peak in metals, and its temperature is correspondingly much higher. Calculation of the basic frequency factor gives 10^{11} sec^{-1}, the entropy factor being neglected. If one adopts a dislocation relaxation model such as that treated by Seeger (1956) the relatively high values of frequency factor and activation energy are consistent with a dislocation having a much higher Peierls force than in metals. The quenching effect would then be explained as temporary pinning of dislocations, for instance by vacancies. Contrary to this explanation is the observation that the peak Δ in the distorted crystal is no higher, despite the increase in dislocation density. Possibly a compensating increase of dislocation interlocking is occurring here.

The frequency factor of the high temperature peak is about 5×10^{13} sec^{-1}, indicating a purely atomic process. The activation energy is lower than those associated with substitutional impurities in silicon (80–110 kcal mole^{-1}), and concentrations of electrically active impurities are too low in the crystals discussed here to explain the observed relaxation strength ($1{\cdot}4 \times 10^{-4}$). There are large concentrations of oxygen in crystals grown from silica crucibles, as these were, and it has recently been suggested that the oxygen atoms are situated interstitially close to two adjacent silicon atoms (Kaiser, Keck and Lange 1956). Motion between adjacent interstitial sites would, in this case, produce a relaxation effect. An activation energy of 50 kcal mole^{-1} is reasonable for this process.

Mullard Research Laboratories,
Cross Oak Lane,
Salfords, Nr. Redhill, Surrey.
15th April 1957.

P. D. SOUTHGATE.

BORDONI, P. G., 1947, *Nuovo Cim.*, **4**, 177.
KAISER, W., KECK, P. H., and LANGE, C. F., 1956, *Phys. Rev.*, **101**, 1264.
SEEGER, A., 1956, *Phil. Mag.*, **1**, 651.

REVIEWS OF BOOKS

Randwertprobleme der Mikrowellenphysik, by F. E. Borgnis and C. H. Papas. Pp. xv + 266. (Berlin: Springer, 1955.) DM 48.

Amongst the many problems which have resulted from the rapid development of microwave physics, diffraction and scattering problems are of primary importance. Very few radar problems are capable of exact analytical solution, so recourse has to be had to approximate methods. But the approximations of optics (very short waves) and of acoustics (long waves) are no longer valid, since in radar the wavelength is of the same order of magnitude as the linear dimensions of the reflector or scatterer.

Many of the approximate methods of dealing with radar problems were developed during the war in the M.I.T. Radiation Laboratory, and one of the most important is Julian Schwinger's variational method. This depends essentially on the fact that the boundary-value problems of microwave physics can be formulated as integral-equation problems. Although the integral equation can be solved only in very special cases, much information can be derived without solving the equation at all. What one does is to express the quantity of physical interest, such as the transmission cross section of the aperture in a plane screen, in a form which is stationary in the sense in which that word is used in the calculus of variations. A brief account of this method will be found in § 6 of C. J. Bouwkamp's article on Diffraction Theory, which appeared in Volume 17 of *Reports on Progress in Physics* (1954).

The book under review contains an account of the analytical formulation of a great variety of reflection and scattering problems which arise in electromagnetic microwave theory and of their approximate solution by Schwinger's variational method. The problems discussed include diffraction by a circular cylinder, a long slit or a long strip, reflection at the junction of two coaxial cables, rectangular wave guides, the approximate determination of the critical frequency of a waveguide of arbitrary cross section, the approximate determination of the characteristic frequencies of a cavity resonator of arbitrary form, and certain aerial problems.

The mathematics of the subject is not easy, and the authors, who have themselves made important contributions in this field, are to be complimented on having written a very good book. E. T. COPSON.

Light Scattering in Physical Chemistry, by K. A. Stacey. Pp. viii + 230. (London: Butterworths Scientific Publications; New York: Academic Press, 1956.) 40*s*.

The fundamental theoretical work of Rayleigh, Mie, Einstein and Gans on the scattering of radiation by particles and by homogeneous media (gases, liquids, solutions) had been completed by 1923. But very little practical use was made of all this until 1944 when, as the author of this book reminds us, " the modern development of light scattering began . . . with Debye's demonstration that the molecular weight of polystyrene could be accurately determined by this method ". The 550 literature references in this book—nearly all after 1943—are evidence of the rapid development that has followed.

The theory is summarized in a chapter of some 60 pages. But even this is too condensed a treatment and one feels that a rather fuller exposition of the fundamentals of the subject would be invaluable in this first book of its kind. For example, the angular distribution of the light scattered by a particle that is not small (larger than 1/20 of a wavelength) is determined by interference effects between the waves scattered from different elements of the particle ; it would make for fuller comprehension of what follows if this were developed carefully instead of the resulting fundamental formulae being merely quoted. The chapter goes on to show how from measurements of transmitted or scattered intensities, the molecular weight of the scattering particles can be found ; and how from the angular distribution of the scattered radiation deductions can be made about the shape of the particle—spheres, rods, random chains and stiffened chains having been thoroughly investigated in the literature.

Then follow chapters on the practice of light scattering applications to high polymers (molecular weights between 2×10^3 and 6×10^8 can be dealt with) and proteins (measurements on tobacco mosaic virus which consists of uniform rods 2700Å long have been very useful in checking the theory). A chapter on poly-electrolytes is particularly interesting to the physicist. When polymers become ionized the repulsion between the charged molecules gives some order to the system and the intensity of scattering falls ; this sort of investigation began only in 1949 when Guinand showed that scattering by polyacrylic acid decreases some 50 times on ionization.

The book is not easy to read or refer to ; this must be so when such a large subject is covered in 230 pages. But more cross references between the sections on theory, practice and applications would help here. And a fuller subject index would be useful ; for example, such topics as coil, chain, rod, ionization and polarization, although treated fully in the text are not mentioned in the index. But while these criticisms affect the ease of use of the book they do not detract from its fundamental usefulness to those who already use light scattering methods and to those who, wondering if they could usefully do so, have hitherto had to search among a mass of original papers. G. F. LOTHIAN.

Electronic Computer, edited by T. E. IVALL. Pp. viii + 167. (London : Iliffe, 1956.) 25*s*.

The book opens with a conventional historic introduction pointing out the significant steps in the gradual evolution of computing aids up to and including the use of electronic techniques. Two of the first electronic digital computers built in this country are mentioned. However a third machine of the same vintage existed at Manchester University and its omission is unfortunate, because as time has proved, it contained a significant contribution to the computer field, namely the cathode ray tube storage system.

A preliminary discussion on the principles of computing and the differences between analogue and digital computers leads on to a more detailed account of both types of machine. However, circuit details precede accounts of the general form of each machine, and many readers will find a difficulty here which would not have been present if the reverse procedure had been adopted.

Applications of these machines to a variety of problems ranging from the training of aircrew to language translation are also discussed. However, in the case of digital machines in particular the reading matter tends to be a list of problems rather than an indication of how to put a problem on to the machine. Omission of this latter detail is serious, since a book of this type should answer the queries of its readers to at least some degree.

The title of the final chapter "Future Machines" does tend to mislead, but valuable information concerning current developments in the computing field is well described there.

Despite these criticisms the book will be useful to anyone wanting to acquire some general knowledge of the computer field. Furthermore, the book contains many photographs illustrating the general form and certain technical details of a variety of existing computers.

D. B. G. EDWARDS.

Lehrbuch der Experimentalphysik, Band III, *Optik und Atomphysik*, Teil 1: *Wellenoptik*, by L. BERGMANN and C. SCHAEFER. Pp. viii + 421. (Berlin: de Gruyter, 1956). DM. 32.

The scope of this book may be indicated by the titles of chapters: (i) Geometrical optics (128 pp.), including discussion of several optical instruments; (ii) Photometry (25 pp); (iii) Dispersion and absorption (72 pp.); (iv) Interference and diffraction (93 pp.), and (v) Polarization and double refraction (88 pp.). Nearly half the book is concerned with ray optics.

The standard of the book is about that corresponding to an honours degree in this country. The author goes rather further than most British University teachers into details of particular cases, and works through some rather tiresome formulae. This method has its advantages. The student may be given courage to tackle the awkward-looking problems that are liable to face the research worker. The disadvantage is that concentration on formulae may obscure general principles.

Although some reference is made to modern work, the book has not been brought up to date in a satisfactory way. For example, four pages are given to measurements of the velocity of light made before 1870, half a page to Michelson's early experiments and nothing at all to the work of Anderson and Bergstrand, though some of the modern determinations of the velocity of waves of radio frequency and of the ratio of the units are mentioned. The book is exceedingly well produced with many excellent diagrams: these alone would make the book deserve the attention of university teachers in this country. The treatment of many parts of the subject is also different in an interesting way from that to which we are accustomed.

R. W. DITCHBURN.

Frequency Modulation, by L. B. ARGUIMBEAU and R. D. STUART. Pp. vii + 96. (London: Methuen; New York: Wiley, 1956.) 8*s*. 6*d*.

Frequency Modulation is a new volume in the series of Methuen's Monographs on Physical Subjects; the authors were formerly on the staff of the Massachusetts Institute of Technology.

The greater part of the book deals with the more elementary aspects of frequency modulation as applied to sound broadcasting. There are chapters on the frequency modulation system, transmitters, receivers, interference, and a brief reference to the television application.

The most important aspect of the subject is the effect of interfering signals on reception conditions; unfortunately, this is dealt with in a somewhat disjointed and confusing manner. The improvement in the signal/receiver-hiss ratio, compared with amplitude modulation, is not placed on a quantitative basis until we have read three-quarters of the book. On the other hand, the contribution to this improvement resulting from pre-emphasis appears much earlier; but this calculation is based on assumptions that would not command general acceptance (American ears apparently have a uniform response up to 21 kc/s!). As a result the importance of pre-emphasis is greatly exaggerated. Impulsive interference is usually important only when the strength of the interference exceeds that of the wanted signal. This case is touched on, but the reader is left in doubt as to the magnitude of the improvement resulting from the use of frequency modulation rather than some other system.

Some basic considerations relating to the frequency modulation system are not mentioned, for instance the factors which determine the peak frequency-deviation, and the distortion of sound programmes in conditions of multipath propagation.

The confusing presentation of the general aspects of the subject is reflected in many individual sections. We frequently find ourselves involved in a mass of detail which tends to obscure the underlying principles. The reader may also get some wrong impressions of the circuits in common use nowadays. The coupled-circuit ratio detector used in the majority of commercial receivers deserves more than a passing reference; again, the crystal oscillator modulated by a reactance valve, which is employed in many British transmitters, is described as ' impractical '.

The Methuen Monographs have established a reputation for succinct and authoritative accounts of modern scientific work. *Frequency Modulation* falls below the standard set by earlier volumes in the series. H. PAGE.

Thermodynamic Functions of Gases, edited by F. DIN. (London : Butterworths Scientific Publications, 1956.) Vol. I, 63*s*., Vol. II, 63*s*.

These are the first two volumes of a work which can reasonably be described as ' monumental '. They form part of a project sponsored by the Department of Scientific and Industrial Research whose aim it is to give a critical survey of all the known thermodynamic properties of some twenty common gases. Vol. I deals with ammonia, carbon monoxide and carbon dioxide; Vol. II with air, acetylene, ethylene, propane and argon. The scope and detail of the work is typified by the article on ammonia which considers the (p, V, T) data for all phases, heat capacities and latent heats, Joule–Kelvin experiments, vapour pressure, the comparison between spectroscopic and calorimetric heat capacities and critical data. The treatment concludes with a (T, S) diagram (which can be obtained also in the form of a wall chart), tabulations of enthalpy, entropy, volume and heat capacity and estimates of the accuracy to which the (T, S) diagram has been constructed. There are some 200 references to earlier surveys and original data. It is clear that the editor's remark—that such work can at times be ' tedious and uninspiring '—is an understatement. (He writes with authority, having prepared three of the eight surveys under review himself as well as contributing an explanatory essay.) The editor and his colleagues are to be congratulated on undertaking a task which will earn the gratitude of a generation of physicists, chemists and engineers. A minor blemish (which 'dates' some of the surveys) is the use of 273·16°K for the ice point instead of the recently accepted value of 273·15°K. A much more serious defect is that the work is—at best—taken up to 1951. It is unlikely that the printing took five years so that the editor must be blamed for failing to revise important parts of the text (e.g. much new data on the condensed phases of argon is ignored in this work.)

R. O. DAVIES.

Physicochemical Calculations, by E. A. GUGGENHEIM and J. E. PRUE. 2nd Edn, Pp. xii+492. (Amsterdam: North-Holland, 1956.) 53*s*.

The first edition of this book was reviewed in *Proc. Phys. Soc.* B, 1955, **68**, 692. The fact that a second edition—with hardly any changes—was needed so soon testifies to its quality. It is perhaps worth remarking that about a quarter of the problems will be found useful by those teaching physics at special honours level. These are the problems on radioactive decay, the basis of kinetic theory, x-rays and structure, dipole moments, heat capacities and the equation of state. A curious error is to be found in problem 7 which (incorrectly) uses both physical and chemical atomic weights.

R. O. DAVIES.

Handbuch der Physik, Vol. 22, *Gas Discharges* II, edited by S. FLÜGGE. Pp. vii+652. (Berlin: Springer, 1956.) DM. 128.

It is sometimes said that scientific handbooks have lost their value, since their contributors act like a group of eminent soloists who form an orchestra and play before an audience without previous rehearsal. This is not quite true. Some authors prefer to present all the available knowledge at the expense of readability, others sacrifice completeness for beauty of presentation or become monographers in a small field. Here none of the extremes is found.

The book starts with a short neatly written article (40 pp.) by F. Llewellyn Jones on ionization growth and breakdown in steady electric fields. It contains experimental facts and ideas about fundamental processes which are active when the breakdown of the gas is gradually approached. 'Growth' is used here in a triple sense: the multiplication of charges from an initial current, the spatial growth and the time growth. There is some emphasis on the work of the author's research group.

Next the glow discharge, mainly at lower pressures, is exhaustively discussed (150 pp.) by G. Francis. Starting with a short section on measurements in discharges and a chapter on similarity laws (which the author has extended in various respects) the observations of the different parameters in the cathode region are recorded, followed by the theory of the cathode fall. Negative glow, Faraday dark space, uniform and striated positive columns and the anode region are treated in detail. The remaining sections deal with cathode sputtering, discharges of special types, those in a magnetic field and in gas mixtures, and finally oscillations and fluctuations, gas pressure differences and separation of gas mixtures. Though there are over 300 references and about an equal number of footnotes, this readable article gives a well-balanced account of the present position.

There is an interesting article by R. G. Fowler on the radiation from low pressure discharges (40 pp.) in which the theory of broadening of lines and associated topics are discussed, as well as the various kinds of excitations by electrons and other particles. This last process still presents some unsolved riddles. The same applies to the absorption of resonance quanta at higher pressures whereas certain types of recombination including the emission processes are well understood. The theory of cascading, collisions of the second kind and the light emission from various discharges concludes the essay.

W. Finkelnburg and H. Maecker partly supported by Th. Peters and G. Buss-Peuckert review comprehensively self-sustaining electric arcs and the thermal plasma (200 pp. in German). The properties of the positive column in various gases are described, such as the potential gradient as a function of pressure and current, spatial distribution of the gas temperature, the forces acting on the column and the means of stabilizing it. There are sections dealing with the theory of the thermal plasma, its thermodynamic treatment, the relations between the transport coefficients of the kinetic theory and the collision cross sections, and the radiation from the plasma. An analysis of the spectra of the column leads to gas temperature, current density and electron concentration. The refractive index of the hot gas is investigated by interference and Schlieren methods, and its motion is discussed, in particular, under the influence of magnetic forces. A short section deals with the cathode and anode mechanism. The excellence of the authors' hobby-horse riding contrasts with a slight scrappiness of the physical aspects in general, and the absence of some fundamentally important material.

L. B. Loeb's chapter (80 pp.) on breakdown with steady and impulse potentials is true to form: it contains much criticism of the work of other researchers as well as that of his own school. Having defined breakdown, the spark and primary and secondary processes, he discusses the growth of Townsend discharges, statistical fluctuations, time lags and breakdown in non-uniform electric fields. The streamer theory in uniform fields is clearly presented, but the reviewer's impression is that a dense cloud of 'Californian' smog still hangs around the goal-posts of our knowledge. A final section is devoted to non-uniform fields, the influence of secondary processes, pulses, electro-negative gases and other topics. The breakdown in high frequency fields (40 pp.) by S. C. Brown is a neat factual report which includes mainly that part of the theory in which diffusion loss is dominant and also the work in the 10 cm wavelength range in uniform and non-uniform fields and with superimposed electric and magnetic fields. It is strikingly incomplete: hardly anything is said about the frequency range below 10^7 c/s and certainly nothing about 'alternating' fields as announced in the title. Many contributions by German and English workers are missing.

The final chapter (50 pp.) is by B. F. J. Schonland on the lightning discharge. The first sections are about the forms and characteristic features of the discharge between clouds and from cloud to ground and the photographic methods and their results. There are also described electrical methods for recording the slow changes of the electric field and its direction, whereas the faster changes are recorded by an oscillograph or a magnetic tape. The measurements of peak currents (from magnetic remanence) and of the time dependence of the current are presented and also the analysis of the field changes in relation to leader and return strokes and the height from which the strokes originate, as well as the results of radar work to elucidate streamer motion in clouds. Finally, the mechanism of breakdown and the propagation of streamers is discussed. A very clear, well-illustrated, up-to-date review.

Altogether this volume is certainly worth while reading. The production is, of course, of the highest standard.

A. VON ENGEL.

Solid State Physics, Vol. 2, edited by F. SEITZ and D. TURNBULL. Pp. xii + 468. (New York: Academic Press; London: Academic Books, 1956.) 80*s*.

This second volume on solid state physics contains five articles. Two of these are concerned with nuclear magnetic resonance. That by C. E. Pake gives an account of the theoretical background of the experimental arrangements used, and of the applications of nuclear resonance in the physics of solids. The survey by W. D. Knight is more restricted in scope, dealing with metals only; it does however cover, in comprehensive detail, the important effects observed in metals.

The applications of neutron diffraction is the subject of a very lucid article by C. G. Shull and E. O. Wollan. The theoretical basis is given briefly, but adequately. About half the article is devoted to the very instructive investigations which have been carried out on crystals which show magnetic effects; other topics dealt with include the location of hydrogen atoms in crystals, and the diffraction of neutrons by liquids.

The theory of lattice vibrations and its application in the theory of the heat capacity of crystals is discussed by J. de Launay. Most of the article is concerned with the vibrations of what are termed ‘central force’ models; this term is somewhat misleading as in none of the cases discussed are the force constants derived from a central force interaction. It is also surprising to find no discussion of ionic lattices. Much attention is devoted to the vibrations of metals, though all treatments so far are rather phenomenological. A point on which clarification would have been welcome concerns the θ (elastic) value for metals; this, according to the calculations of Dr. de Launay, appears to have a different value in different theories even though the elastic constants are the same. The comparison between theory and experiment is given in any detail only for the diamond type crystals. One would have welcomed the inclusion of the theoretical and the experimental results for rocksalt and sylvine.

The effects on solids of irradiation by neutrons and by light charged particles is discussed by F. Seitz and J. S. Koehler. A comprehensive treatment is given which includes the elementary processes, the production of displacement and of temperature spikes, and the generation of defects and of dislocations. The experimental observations on crystals with various types of bonding are discussed in some detail, i.e. the change of volume in the case of non-metals and the change in resistance in the case of metals. The effects on the heat conductivity appear to have been omitted.

The book will be indispensable to physicists working on those aspects of physics which it covers. It can also be warmly recommend to all physicists interested in the rapidly growing subject of solid state physics.

M. B.

Elementary Crystallography (An Introduction to the Fundamental Geometrical Features of Crystals) by M. J. Buerger. Pp. xxiii + 528. (New York: Wiley; London: Chapman and Hall, 1956.) 70*s*.

Crystallography is not a difficult subject, but it cannot be learnt without effort. Students will therefore welcome this new approach, at an elementary level, and in modern symbolism, by a distinguished American crystallographer. He writes in the preface—" Nearly every mineralogy, metallurgy, chemistry, and physics textbook now contains a short section on crystallography. These abbreviated treatments can hardly be expected to present more than some of the conclusions of the science. Something was plainly needed to explain the reasoning leading to these conclusions. This book was written to supply a fuller account of the subject." The book may be roughly divided into two parts. First, the principles of symmetry operations are developed, and applied to derive the 32 point groups; these are enumerated and classified, and illustrated with copious examples provided by the various crystal forms. The latter half of the book presents a systematic account of the 230 space groups, supported by some 200 carefully thought out, and beautifully prepared diagrams. Numerous other topics of crystallographic interest are touched upon, in particular there are three final chapters on abstract group theory. The writing of this book must have involved immense labour, but there can be no doubt that it will prove to have been worth while. M. A. JASWON.

BOOK NOTICES

Engineering Analysis: A Survey of Numerical Procedures, by S. H. Crandall. Pp. ix + 417. (New York: McGraw-Hill, 1956.) 71*s*. 6*d*.

Calcul numérique, calcul physique. Application a la thermocinétique, by V. Vernotte. Pp. 344. (Paris: Publication Scientifiques et Techniques du Ministère de l'Air, No. 319, 1956.) 2200 *fr*.

Étude théorique et expérimentals de la stabilité des chambres d'équilibre situées en aval d'une galerie en charge alimentée par un canal a écoulement libre, by S. Gerber. Pp. 122. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air, No. 320, 1956.) 1400 *fr*.

Photographic Sensitivity, Vol. I. *Hakone Symposium*. Proceedings of as Symposium held at the Lake Hakone, Japan, in September 1953. Edited by Shin Fujisawa. Pp. 3 + 147. (Tokyo: Maruzen Co. Ltd., 1956.) $4.00.

Steric Effects in Organic Chemistry, edited by M. S. Newman, Pp. vii + 710. (New York: Wiley, 1956.) 100*s*.

The Constitutional Diagrams of Alloys: A Bibliography. Originally compiled by J. L. Haughton; second edition compiled by A. Prince. Pp. 324. (London: Institute of Metals, Monograph and Report Series, No. 2, 1956.) 35*s*.

Electrical Energy, edited by H. G. Foster. (London: Morgan Bros.) Vol. 1, No. 1, September 1956, monthly. 3*s*. per issue.

Optical Design with Digital Computers (Proceedings of a Symposium held in the Physics Department, Imperial College, on 5th, 6th and 7th June, 1956) Pp. viii + 65. (London: Imperial College, 1956). 10*s*.

A limited number of copies of this report is available, on application to Professor W. D. Wright, Technical Optics Section, Physics Department, Imperial College, London, S.W.7. Cheques should be made out to Mr. W. Thompson, Physics Department, Imperial College.

Metallurgy of the Rarer Metals. No. 5. *Molybdenum*, by L. Northcott. Pp. xii + 222. (London: Butterworths Scientific Publications, 1956). 40*s*.

Algebraic Varieties, by M. Baldassarri. Pp. ix + 195. (Berlin: Springer, 1956). (Ergebnisse der Mathematik und ihrer Grenzgebiete, Neue Folge, Heft 12). 36 DM.

Abelsche Funktionen und algebraische Geometrie, by F. Conforto. Pp. xi + 276. (Berlin: Springer, 1956). 41.80 DM.

Constantes sélectionnées pouvoir rotatoire naturel. I. *Stéroïdes*, by J.-P. Mathieu and A. Petit. Pp. 507. (Paris: Masson, 1956).

PARIS PUBLICATIONS SCIENTIFIQUES ET TECHNIQUES DU MINISTÈRE DE L'AIR:

Étude de la convection par les fils aux faibles nombres de Reynolds, by J. Gosse. No. 322, 1956. 1200 fr.

Résolution de divers problèmes du type Stokes–Beltrami posés par la technique aéronautique, by P. Brousse. No. 323, 1956. 1200 fr.

Contribution a l'étude des propriétés et des modes de décomposition de la phase protoxyde de fer, by R. Collongues. No. 324, 1957. 1500 fr.

Application de la micrographie électronique a l'étude de certains aspects de la surface des cristaux et de la nature des joints intergranulaires dans la cas de l'aluminium, by P. Bussy. No. 325, 1957. 1200 fr.

Séminaire d'aérothermique de la Faculté des Sciences de Paris, Année 1954–55. No. N.T. 63, 1957. 1200 fr.

CONTENTS OF SECTION A

PAGE

Dr. A. M. Lane and Mr. J. E. Lynn. Fast Neutron Capture below 1 Mev: The Cross Sections for ^{238}U and ^{232}Th 557

Mr. J. W. Stephenson and Mr. K. Smith. Tabulation of the Wigner 9*j* Coefficients 571

Dr. W. D. Allen and Dr. A. T. G. Ferguson. The Fission Cross Sections of ^{233}U, ^{235}U, ^{238}U and ^{239}Pu for Neutrons in the Energy Range 0·030 Mev to 3·0 Mev 573

Mr. A. W. Dalton, Mr. S. Hinds and Dr. G. Parry. An Investigation of the Reaction ^{31}P(d, p)^{32}P by Magnetic Analysis 586

Dr. R. McWeeny. Electronic Absorption Spectra of Heterocyclic Systems. I: A Comparison of Theoretical Methods: Pyridine and Pyrazine . . 593

Dr. J. C. Barton, Mr. D. A. Campbell and Mr. R. C. Read. An Analogue Method for Studying Multiple Scattering 605

Research Notes:

Dr. E. W. Elcock. A Note on the Ising Model with Second Nearest Neighbour Interactions 615

Mr. A. M. Arthurs, Mr. R. A. B. Bond and Mr. J. Hyslop. The Oscillator Strengths of the $1s\sigma$–$2p\sigma$, $2p\sigma$–$2s\sigma$, $2p\sigma$–$3d\sigma$ and $2p\sigma$–$2p\pi$ Transitions of HeH^{2+} 617

Dr. M. J. Seaton. The Use of Extrapolated Quantum Defects as a Check on Calculated Phases for Scattering of Electrons by Positive Ions . . 620

Letters to the Editor:

Dr. R. F. Barrow. On the Rotational Analysis of Bands of the $A^3\Pi_{0^+}$–$X^1\Sigma^+$ System of TlCl 622

Reviews of Books 624

Book Notices 627

Contents of Section B 628

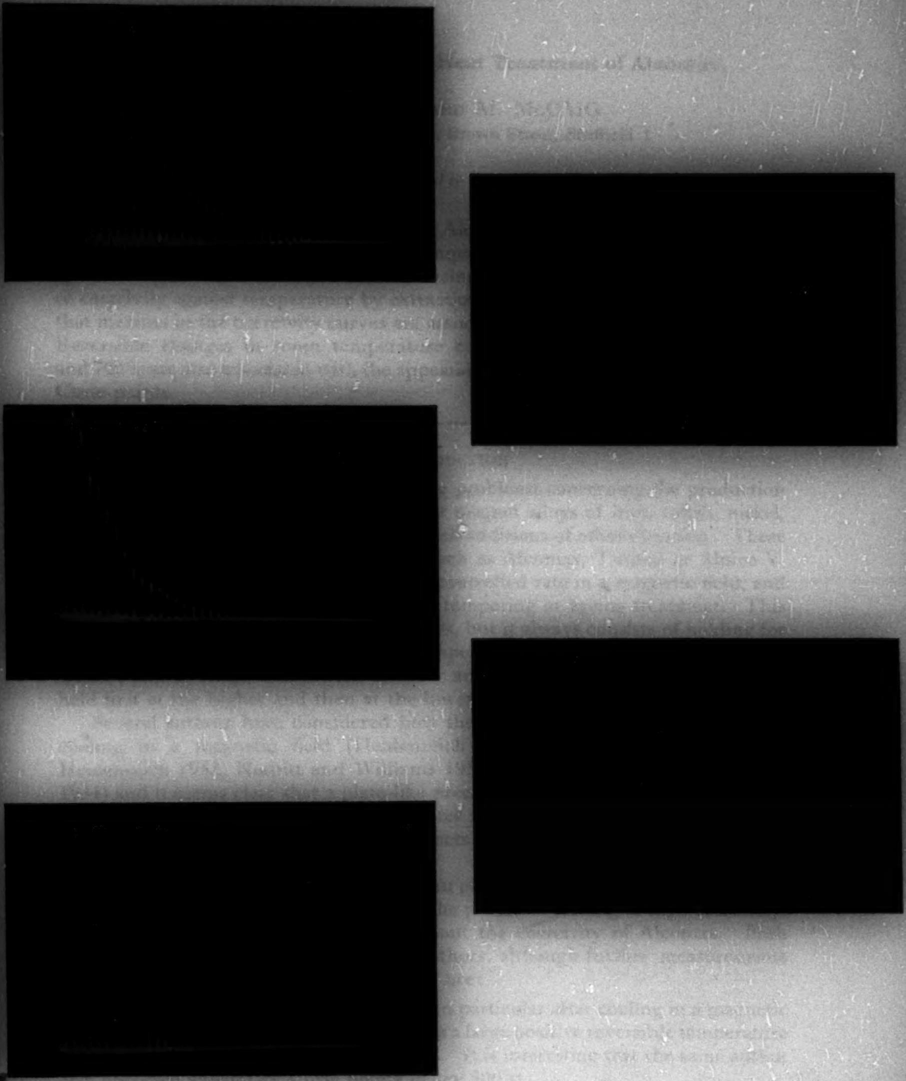


Figure 4.

Processes occurring during the Heat Treatment of Alcomax

By A. G. CLEGG and M. McCAIG
Permanent Magnet Association, Brown Street, Sheffield 1

MS. received 8th April 1957, *and in final form 6th June* 1957

Abstract. The permanent magnet alloy Alcomax has been investigated after different heat treatments using two techniques. Curves of saturation intensity against temperature have been obtained using a Sucksmith balance, and curves of coercivity against temperature by extraction from a solenoid. It is concluded that maxima in the coercivity curves are associated with subsidiary Curie points. Reversible changes in room temperature coercivity produced between 600°c and 700°c are also associated with the appearance and disappearance of subsidiary Curie points.

§ 1. Introduction

There are a number of outstanding problems concerning the production of high coercivity in the permanent magnet alloys of iron, cobalt, nickel, aluminium, copper and optional small additions of other elements. These alloys, which have various trade names such as Alcomax, Ticonal or Alnico V, are first made anisotropic by cooling at a controlled rate in a magnetic field, and then given a high coercivity by a suitable tempering or ageing treatment. This treatment can be conducted in several ways, but it always consists of holding for one or two hours to several days at temperatures between 650°c and 550°c. Often a pair of temperatures such as 590°c and 560°c is used, the magnets being held first at the higher and then at the lower temperature.

Several authors have considered how the magnets are made anisotropic by cooling in a magnetic field (Heidenreich and Nesbitt 1952, Nesbitt and Heidenreich 1952, Nesbitt and Williams 1955, Kronenberg 1954, Fahlenbrach 1954) and it seems clear that a plate-like precipitate is formed. The plates form in certain crystallographic planes, but plates nearly normal to the magnetic field are inhibited. The anisotropy of the magnets is attributed to the shape anisotropy of these oriented plate-like precipitates.

The anisotropy can be produced without giving the alloy a very high coercivity and this paper is concerned mainly with the processes leading to high coercivity.

There are two extraordinary facts about the coercivity of Alcomax. Both have been previously reported by other authors, although further measurements are described in this paper. These facts are:

A. After certain heat treatments, and in particular after cooling in a magnetic field but before the final tempering, there is a large positive reversible temperature coefficient of coercivity (Jellinghaus 1943). It is interesting that the same author also observed subsidiary Curie points below 500°c.

B. If the fully tempered alloy is re-heated to a higher temperature in the range 650°c to 750°c, the coercivity at room temperature is reduced; if the alloy is then re-tempered below 600°c the coercivity is partly or wholly restored (Hansen 1955, van der Steeg and de Vos 1956).

In this investigation saturation magnetization and coercivity have been measured as a function of temperature after various heat treatments.

§ 2. Experimental Details

Alcomax III of the following nominal composition was used: 25% Co, 13·5% Ni, 7·8% Al, 3% Cu, 0·8% Nb and balance Fe. The alloy was cast in blocks with dimensions 38 mm × 35 mm × 10 mm. After heat treatment the blocks were cut with a slitting wheel into bars for the coercivity measurements, and also into small samples, which could be accommodated in a holder 10 mm long and 3 mm in diameter, for saturation magnetization measurements, which were carried out at various temperatures using a ring balance (Sucksmith 1939). The estimated effective field used was about 9000 oersteds. For measurements of coercivity at different temperatures a small tube furnace which was non-inductively wound was placed inside a solenoid capable of producing fields up to 3000 oersteds. A search coil with 50 turns, connected to a ballistic galvanometer, was fitted inside the furnace. The sample could be moved in and out of this coil. The sample was magnetized and then the demagnetizing field found such that no deflection of the galvanometer resulted from moving the sample. This field is actually the intrinsic coercivity ${}_{\mathrm{J}}H_{\mathrm{c}}$ for the sample; for Alcomax it differs only slightly from ${}_{\mathrm{B}}H_{\mathrm{c}}$.

§ 3. Results

In figure 1, top portion, (σ, T) curves relate to the alloy after the following treatments: *a*, oil quenched from 1250°c; *b*, cooled in a magnetic field from 1250°c to 600°c at 1·2 deg sec^{-1}; *c*, tempered for 48 hours at 590°c followed by 48 hours at 560°c; *d*, tempered for 48 hours at 650°c and *e*, tempered for 4 hours at 700°c. The last three treatments were all preceded by cooling in a magnetic field from 1250°c, so that sample *c* should have the normal Alcomax properties. Note the vertical displacement of curves *b*–*e*. In the bottom left-hand corner of the same figure (coercivity, temperature) curves are shown for samples with heat treatments *b*, *c*, *d* and *e*. The coercivity of the quenched sample was less than 10 at room temperature and a curve is not included for this state. The same abscissa scale is used for all curves on this figure, and it should be noted that maxima in the coercivity curves occur at about the same temperature as kinks indicating subsidiary Curie points of the (σ, T) curves.

The arrangement of figure 2 is similar, but it is drawn to illustrate the reversible changes in the (coercivity, temperature) curves and Curie points produced by successive treatments at 700°c and 590°c. The (σ, T) and (H_{c}, T) curves after the first and second treatments at 700°c are very similar, being little affected by the intermediate treatment at 590°c. The curves after the 590°c treatment are similar to those for the fully treated state in figure 1. Tempering at any temperature between 590°c and 700°c produces (σ, T) and (H_{c}, T) curves typical of that temperature more or less independently of previous treatments.

Reversible changes in remanence B_{r}, energy value $(BH)_{\max}$ and coercivity H_{c} after spoiling at various temperatures up to 765°c and re-tempering for 48 hours at 590°c plus 48 hours at 560°c have also been investigated and the results are shown in the table.

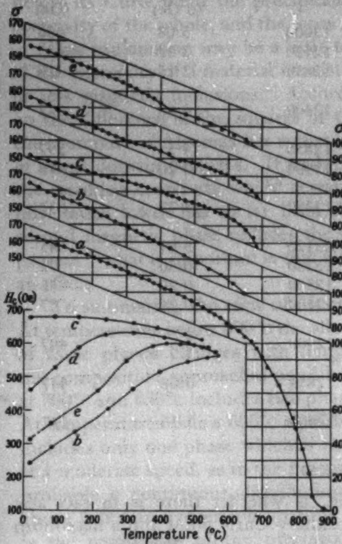


Figure 1.

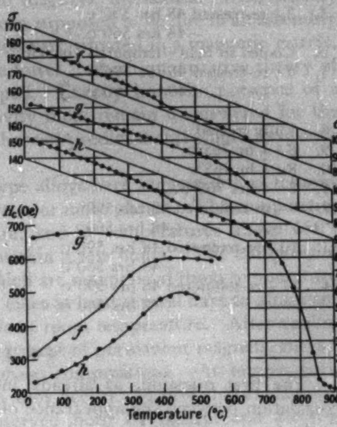


Figure 2.

Saturation magnetization per unit mass and coercivity plotted against temperature.

Figure 1. *a*, quenched 1250°c.
b, cooled 1·2 deg sec^{-1} from 1250°c.
c, tempered 48 hrs 590°c, 48 hrs 560°c.
d, tempered 48 hrs 650°c.
e, tempered 4 hrs 700°c.

Figure 2. *f*, cooled 1·2 deg sec^{-1} from 1250°c, tempered 4 hrs at 700°c.
g, *f* and tempered 16 hrs at 590°c.
h, *g* and tempered 4 hrs at 700°c.

At 650°c the magnetic properties reach steady values after 1 hour and are unaffected by a further 14 hours at 650°c. After 1 hour at 650°c the magnetic properties are substantially independent of any previous tempering at a lower temperature, compare examples 2 and 6 in the table. Re-tempering at 590°c and 560°c restores at least 90% of the properties of any of the samples spoiled at 650°c, compare examples 1 and 7. At 765°c the spoiling is not only more severe but increases and gradually becomes more permanent with time. After half an hour at 765°c, 70% of the $(BH)_{\max}$ can be restored by re-tempering, but after 4 hours only 30% can be restored, compare examples 8, 12 and 14.

Similar experimental results have been obtained with Alcomax II, another anisotropic alloy with slightly different composition. These results showed that the behaviour at 700°c is intermediate between that at 650°c and 765°c.

Heat Treatment	B_r(G)	$(BH)_{\max} \times 10^{-6}$ (G Oe)	H_c (Oe)
1. Fully treated	13000	5·08	638
2. Fully treated + 1 hr 650°c	12920	3·30	435
3. Fully treated + 15 hrs 650°c	12850	3·15	417
4. 2 + re-treated 48 hrs 590°c + 48 hrs 560°c	12400	4·68	632
5. 3 + re-treated 48 hrs 590°c + 48 hrs 560°c	12550	4·5	597
6. Cooled in field + tempered 1hr 650°c	12900	3·10	418
7. 6 + re-treated 48 hrs 590°c + 48 hrs 560°c	12350	4·73	637
8. Fully treated	13120	5·01	645
9. 8 + ½ hr 765°c	13380	0·50	72
10. 8 + 1 hr 765°c	12500	0·29	53
11. 8 + 4 hrs 765°c	13100	0·26	47
12. 9 + re-tempered 48 hrs 590°c + 48 hrs 560°c	12900	3·5	497
13. 10 + re-tempered 48 hrs 590°c + 48 hrs 560°c	12850	3·08	447
14. 11 + re-tempered 48 hrs 590°c + 48 hrs 560°c	12000	1·56	247

§ 4. DISCUSSION

The first question, as already indicated, is whether there is in fact any correlation between the appearance of a subsidiary Curie point and a maximum in the (coercivity, temperature) curve. Samples tempered at 700°c have a Curie point at about 450°c which coincides very closely with the maximum in coercivity. Samples tempered at 650°c have indications of Curie points at about 300°c and 450°c, while the (coercivity, temperature) curve has a flat-topped maximum approximately covering this temperature range. After the double 590°c + 560°c tempering treatments, no Curie points are observed and the (coercivity, temperature) curve has no maximum above room temperature. Elsewhere (Clegg 1955) it has been shown that the coercivity does decrease at temperatures below room temperature.

The behaviour of the sample cooled from 1250°c is less clear (figure 1, curve *b*). The average rate of cooling from 1250°c to 600°c is 1·2 deg sec^{-1}; in the important region between 900°c and 750°c it is about 1 deg sec^{-1}, and for the last 150°c, 0·7 deg sec^{-1}. The slope of the (intensity, temperature) curve is steeper than that for the sample quenched from 1250°c, and is compatible with the assumption that the sample contains a continuous variation of compositions with different Curie points. If there is in fact a particular composition with its own Curie point which would precipitate out at each temperature, it is not impossible that such a continuous distribution of composition might result from a steady rate of cooling. The corresponding (coercivity, temperature) curve is still rising at 500°c, which according to the argument below, suggests that some parts of the precipitate should have Curie points higher than this temperature, where they might be difficult to observe.

The original suggestion by Jellinghaus was that the precipitate had a low coercivity and the matrix a high coercivity. Above the Curie point of the precipitate, the matrix exerts its full coercivity which falls slowly with temperature.

Below its Curie point the precipitate, acting as a magnetic shunt, reduces the coercivity of the whole, and the more so the greater its intensity of magnetization.

This explanation may be a little too simple. Néel (1946) published a theory of the coercivity of a material consisting of a magnetic phase with non-magnetic or less magnetic inclusions. According to this theory the coercivity depends on the difference of the squares of the two magnetic intensities. Alternatively the coercivity of Alcomax has sometimes been supposed to be due to an assembly of single domain particles. If such particles were situated in a magnetic matrix, the resulting coercivity would certainly be lower, but no quantitative theory of how much lower has so far been developed. Exactly how the presence of a second magnetic phase reduces the coercivity must remain unanswered for the present; what is important is that all the theories predict that it would have such an effect.

To summarize, the view of Alcomax type alloys here suggested is as follows. At temperatures below 800°c the alloy has two or more phases. The composition of these phases changes with temperature, and at temperatures above 650°c the composition approaches equilibrium within a few hours. The composition at 700°c and 650°c includes two phases which are magnetic at room temperature. At temperatures below 600°c equilibrium takes at least several days to attain and includes only one phase which is magnetic at room temperature. After cooling at a moderate speed, as in the practical treatment of permanent magnets, there is probably a continuous distribution of phase compositions. At temperatures above 700°c an irreversible spoiling of the alloy occurs. A probable explanation is the coarsening of the precipitate observed both by optical and electron microscopes. Both the single domain particle theory and the inclusion theory of Néel indicate that this coarsening of the precipitate should reduce the coercivity.

Although changes in the composition of the precipitate, as indicated by changes in the room temperature coercivity, occur fairly quickly at some temperatures, this is not the case in a sample quenched from over 900°c in which the coercivity is very low and rises only very slowly on tempering between 550°c and 700°c (McCaig 1957).

In the practical manufacture of Alcomax magnets, the process of cooling in a magnetic field may be started at about 1250°c or 950°c, but not at intermediate temperatures at which spoiling occurs. The final properties of the magnets after tempering are very similar whether the 1250°c or 950°c temperature is used. If the magnets are tested at the intermediate stage, after field cooling but before tempering, those cooled from 950°c have about double the coercivity of those cooled from 1250°c (say 300 to 400 oersteds instead of 150 to 200 oersteds). This difference occurs although the actual rate of cooling below 900°c is the same in each case. Evidently something happens to the alloy while it is held for 10 or 20 minutes at 950°c before field cooling, which does not take place if it cools fairly quickly through this temperature; possibly the precipitate formed at lower temperatures can be nucleated at 950°c.

ACKNOWLEDGMENTS

The authors thank the Permanent Magnet Association for permission to publish this paper and Mr. J. E. Gould for helpful interest and advice.

References

Clegg, A. G., 1955, *Brit. J. Appl. Phys.*, **6**, 120.

Fahlenbrach, H., 1954, *Tech. Mitt. Krupp*, **12**, 177.

Hansen, J. R., 1955, *Conference on Magnetism and Magnetic Materials, Amer. Inst. Elect. Engrs., Pittsburgh*, p. 198.

Heidenreich, D. R., and Nesbitt, E. A., 1952, *J. Appl. Phys.*, **23**, 352.

Jellinghaus, W., 1943, *Arch. Eisenhuttenw.*, **16**, 247.

Kronenberg, K. J., 1954, *Z. Metallk.*, **45**, 440.

McCaig, M., 1957, *Proc. Phys. Soc.* B, 823.

Néel, L., 1946, *Annales de l'Université de Grenoble*, **22**, 299.

Nesbitt, E. A., and Heidenreich, R. D., 1952, *Elect. Eng.*, **71**, 530.

Nesbitt, E. A., and Williams, H. J., 1955, *J. Appl. Phys.*, **26**, 1217.

van der Steeg, M. G., and de Vos, K. J., 1956, *J. Appl. Phys.*, **27**, 1250.

Sucksmith, W., 1939, *Proc. Roy. Soc.* A, **170**, 551.

The Unusual Magnetic Properties of Quenched Alcomax III

By M. McCAIG

Permanent Magnet Association, 84 Brown St., Sheffield 1

MS. received 28th May 1957

Abstract. If the permanent magnet alloy Alcomax III is quenched from a high temperature instead of the usual treatment it has the properties of a rather unusual soft magnetic material. The permeability is almost constant up to flux densities over 5000 gauss and the ratio of remanence to saturation intensity is less than 0·1. Details of these properties are given and their theoretical implications considered

§ 1. INTRODUCTION

ALCOMAX III is an alloy of nominal composition 24% Co, 13·5% Ni, 8·0% Al, 3% Cu, 0·8% Nb and balance Fe. It is made into an anisotropic permanent magnet alloy by cooling at a controlled speed in a magnetic field, applied in the direction along which it is afterwards to be magnetized, followed by a prolonged tempering at about 600°c. Hoselitz and McCaig (1951) mentioned that if this alloy is quenched from a high temperature it has quite different properties including permeability that varies little with induction over a wide range, low remanence and low coercivity. The remanence may be as low as one tenth of the saturation intensity.

These properties are sufficiently unusual to be worth reporting for their own sake. Although there does not appear to be any demand for them at present, the knowledge that they exist might suggest an application. The purpose for which this work was initiated was to understand better the processes occurring in Alcomax.

The apparatus used in this work was quite standard. The quenched Alcomax was either in the form of anchor rings or rectangular test pieces. The former were tested by the usual ballistic method; the latter, which were used in an experiment in which the coercivity was increased by tempering, were tested with the B.S.I. Faraday disc apparatus.

§ 2. RESULTS

A general idea of the properties of quenched Alcomax III is given by figure 1 and the table. The initial permeability μ_i is not much less than the maximum permeability μ_{max}. In some cases the results show the two permeabilities equal, but it is possible that measurements at fields smaller than 1 oersted would reveal a slightly lower initial permeability. It is however certainly possible to obtain permeabilities that are almost constant up to a flux density of 5000 gauss.

	μ_i	μ_{max}	B_{max} (G)	B_r (G)	H_c (Oe)	W_h (erg cm^{-3} cycle^{-1})
Oil quenched 950°c	500	500	12000	1450	2·0	2870
Water quenched 950°c	385	385	13350	1000	2·0	5630
Oil quenched 1250°c	167	490	10000	1600	2·0	6950

The coercivities H_c and energy losses per cycle W_h are not much greater than those of many commercial soft magnetic materials. The remanence B_r lies between 1000 and 1600 gauss; these are remarkably low values for a material with a saturation $4\pi J_s$ of about 14 000. The figures for B_{max} in the table are the maximum inductions for the hysteresis curves from which B_r, H_c and W_h were deduced. It was not possible to saturate these anchor rings completely.

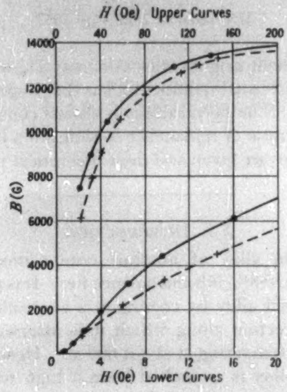


Figure 1. Initial magnetization curves. Full curve oil quenched 1250°c. Broken curve water quenched 950°c.

The variation of W_h with B_{max} is shown in figure 2 for the anchor ring quenched in oil from 1250°c. The hysteresis loops are thin but elongated and a considerable proportion of the area is in the region between 50 and 100 oersteds.

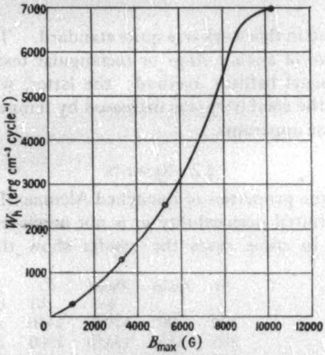


Figure 2. Hysteresis loss per cycle as a function of B_{max} for a sample.

These properties are unique. The nearest parallel in commercial materials is Perminvar. If some use should arise for such properties there is ample scope to seek improvements by varying the composition or heat treatment. At present the alloy could be produced in solid or powder form but not in sheets.

A test piece of Alcomax which had been quenched and had a coercivity of 1·5 oersteds was tempered at 590°c. After 1500 hours the coercivity had risen to 150 oersteds and was still rising at about 5 oersteds per 100 hours. Further test pieces were water quenched from 1250°c and given a variety of heat treatments, and the course of their recovery of remanence and coercivity is shown in figure 3. The remanence reaches values around 8000 gauss, which are normal for Alcomax which has not been field treated, but the coercivity could not be raised above 250 oersteds.

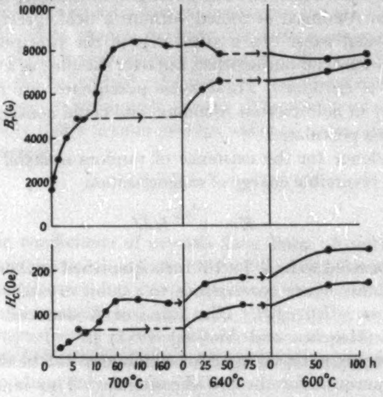
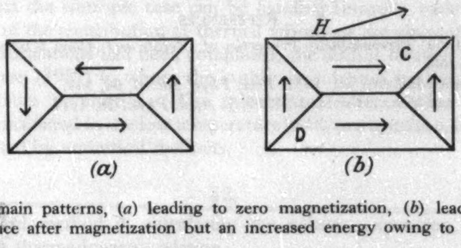


Figure 3. Remanence and coercivity of Alcomax III water quenched from 1250°c and tempered for various times at 700°c, 640°c and 600°c. The broken arrows indicate that the test piece was transferred to the furnace at the next lower temperature.

§ 3. Discussion

Hoselitz and McCaig (1951) mentioned a tentative explanation of the low remanence values of quenched Alcomax. As the alloy cools through its Curie point it takes up a domain pattern resulting in zero magnetization such as that shown in figure 4(*a*). Suppose that the local directions of magnetization become



(*a*) (*b*)

Figure 4. Domain patterns, (*a*) leading to zero magnetization, (*b*) leading to normal remanence after magnetization but an increased energy owing to free poles at C and D.

local preferred directions. Let a magnetizing field be applied in direction H and removed: figure 4(*b*). If the magnetization returned everywhere to the easy direction but with a positive component in the direction in which H had been applied the remanence would have a normal value. Figure 4(*b*) shows that this arrangement would involve an increase in energy due to the presence of free poles at C and D. If the domains are free to return to the original directions in figure 4(*a*), the energy would be reduced and a low remanence would result. Presumably this domain movement occurs in a quenched alloy but is inhibited in a slowly cooled or tempered alloy.

There is now independent evidence for this explanation. Nesbitt and Williams (1955) have produced electron micrographs showing that when Alnico V, which is similar to Alcomax, is cooled without a field, precipitates are formed parallel to the crystal axes. In a small region the precipitates appear to be parallel to the same axis and one another, but over the alloy as a whole the direction of the precipitate is random. These same precipitates are responsible for the uniaxial anisotropy in field treated Alcomax, and could consequently account for the domain patterns postulated.

Additional evidence for the existence of random uniaxial anisotropy can be obtained from the reversible energy of magnetization

$$E = \int_{J_r}^{J_s} H dJ.$$

E has been measured to be $7{\cdot}3 \times 10^4$ for a quenched anchor ring. According to Hoselitz (1952) this figure corresponds to a cubic crystal anisotropy constant K of $+5{\cdot}5 \times 10^5$ or $-3{\cdot}6 \times 10^5$. The value of K deduced from a quenched columnar sample (Hoselitz and McCaig 1951) is $+1 \times 10^5$. There is thus considerable anisotropy not accounted for which may well be the random uniaxial term required to account for the low remanence. This random anisotropy is however much smaller than the uniaxial anisotropy in the field treated alloy.

The sluggish response to tempering suggests a lack of nuclei in the quenched alloys. When the alloy is cooled at the optimum speed it is probable that nuclei are formed at temperatures between 900°c and 700°c which enable the precipitate to grow at lower temperatures.

ACKNOWLEDGMENTS

The author thanks the Permanent Magnet Association for permission to publish this paper.

REFERENCES

HOSELITZ, K., 1952, *Ferromagnetic Properties of Metals and Alloys* (Oxford : Clarendon Press), p. 66.
HOSELITZ, K., and MCCAIG, M., 1951, *Proc. Phys. Soc.* B, **64**, 549.
NESBITT, E. A., and WILLIAMS, H. J., 1955, *J. Appl. Phys.*, **26**, 1217.

On the Thermal Expansion of Solids

By M. BLACKMAN

Physics Department, Imperial College, London

MS. received 3rd June 1957

Abstract. The thermal expansion of alkali halides has been investigated at very low temperatures and at high temperatures. Attention has been concentrated on the isotropic cases, for which a formal expression for the Grüneisen constant has been obtained. It is shown that the Grüneisen constant is expected to increase somewhat as the temperature is increased. It is further shown that the contributions from the individual vibrations to the Grüneisen constant at low temperatures vary over a very wide range. The Grüneisen rule, however, is shown to require only that a certain average value of these contributions should not vary with frequency.

§ 1.

The expansion coefficients of crystals have been recently the subject of a number of experimental and theoretical investigations. While most of the experimental work has been concerned with metals, the theoretical considerations have been concerned with cases where the forces between particles can be handled in a satisfactory manner; this confines, at present, the calculations to models where the forces between any two particles are represented by means of a potential depending only on the distance between the particles. Barron (1955) has published such an investigation for the face centred cubic crystals of the argon type. For alkali halides there exists an early calculation by Born (1923). This was confined to cases where the elastic anisotropy was small. Though the details of the calculation are rather obscure, Born was the first to point out that the Grüneisen constant γ did in fact depend on the temperature of the crystal. The value for low temperatures γ_0 was about 30% smaller than γ_∞ the value at high temperatures. A variation in the same direction, though smaller in magnitude, was found by Barron for the argon-type face-centred lattice.

In investigations described below, the ionic crystals of the rocksalt type are studied. A much simpler method than that used by Born has been adopted, and it is shown that the isotropic case can be handled formally without difficulty. Other aspects of the contribution of thermal vibrations are also considered.

After the calculations had been completed, the author learned of an investigation by Barron (1957) to whom the author is indebted for sending him the manuscript before publication. The calculations differ in that the isotropic case was not considered in the low temperature part, two special cases of γ_0 being instead evaluated by numerical methods.

§ 2.

The standard method of calculating the volume expansion coefficient β depends on the thermodynamic relation

$$\frac{\beta}{\kappa}=-\frac{\partial^2 F}{\partial V\partial T} \qquad \ldots\ldots(1)$$

where κ is the isothermal compressibility, F the free energy, V the volume and T the absolute temperature.

Assuming that the temperature dependent part of free energy of the crystal is entirely vibrational in character, the free energy can be written as

$$F=U_0+\sum_\nu kT\log\{(1-\exp(-h\nu/kT)\}. \qquad \ldots\ldots(2)$$

Here U_0 is the sum of the internal energy and the zero point energy of the crystal. The summation is taken over all the normal vibration of the crystal. From (1) it follows that

$$\frac{V\beta}{\kappa}=\sum_\nu \gamma_\nu E\left(\frac{h\nu}{kT}\right) \qquad \ldots\ldots(3)$$

where E is the heat capacity of a linear oscillator and

$$\gamma_\nu=-\frac{V}{\nu}\left(\frac{\partial\nu}{\partial V}\right)_T. \qquad \ldots\ldots(4)$$

This formula can be linked with the usual theoretical definition of Grüneisen's constant γ by the relation

$$\gamma=\frac{V\beta}{\kappa C_V}=\frac{\sum_\nu \gamma_\nu E(h\nu/kT)}{\sum_\nu E(h\nu/kT)}. \qquad \ldots\ldots(5)$$

The quantity γ so defined is a parameter which in principle will vary with the temperature.

The Grüneisen relation in its strict sense, namely that γ is independent of temperature, will follow if all γ_ν are independent of frequency. This would certainly not be expected in a lattice and, as will be shown below, γ_ν varies over a very large range even if special vibrations are considered.

It is however an empirical fact that in many cases the experiments do not show any marked variation of γ with temperature. A much less stringent theoretical condition should therefore exist. Such a condition can be seen if one rewrites (5) in the form

$$\gamma=\frac{\int \bar{\gamma}_\nu\rho(\nu)E(h\nu/kT)d\nu}{\int \rho(\nu)E(h\nu/kT)d\nu}. \qquad \ldots\ldots(6)$$

Here a new quantity $\overline{\gamma_\nu}$ has been introduced. This is defined by considering two contours (in reciprocal space), the frequencies associated with these being ν and $\nu+\Delta\nu$ respectively. An arithmetic average value of γ_ν is found for all vibrations lying between ν and $\nu+\Delta\nu$. This value we term $\overline{\gamma_\nu}$. It then follows from (6) that if

$$\overline{\gamma_\nu}=\text{constant} \qquad \ldots\ldots(7)$$

then $\gamma=\overline{\gamma_\nu}$, and this is then independent of temperature in the theory used. This condition seems to be the theoretical requirement for the Grüneisen rule.

§3.

In the general theory, there are two special cases in which (6) reduces to an especially simple form (Barron 1955). At sufficiently high temperatures

($\gamma = \gamma_\infty$) the heat capacity of a linear oscillator becomes constant, i.e.

$$\gamma_\infty = \int \overline{\gamma_\nu} \rho(\nu) d\nu \Big/ \int \rho(\nu) d\nu \qquad \ldots\ldots (8)$$

which represents an arithmetic average over all γ_ν. The second case refers to low temperatures, i.e. the region where the T^3 law for the heat capacity of the crystal holds with sufficient approximation. The γ_0 value, which is most easily derived directly from the explicit form of the free energy in this case, takes the form

$$\gamma_0 = \left\{ \iiint \sum_{j=1}^{3} c_j^{-3} \gamma_j^{(C)} d\Omega \Big/ \iiint \sum_{j=1}^{3} c_j^{-3} d\Omega \right\} \qquad \ldots\ldots (9)$$

where

$$\gamma_j^{(C)} = \frac{1}{3} - \frac{V}{c_j}\frac{dc_j}{dV} = \frac{1}{3} - \frac{r}{6c_j^2}\frac{dc_j^2}{dr}. \qquad \ldots\ldots (10)$$

Here c_j is the velocity of a wave, in a given direction, in one of the three acoustic branches, and r is the lattice spacing, or any quantity proportional to it.

§4.

The velocity of the elastic waves which are needed in the above calculation will involve a knowledge of the elastic constants. Since we are interested in the variation with spacing it is essential to obtain these constants without making any use of the equilibrium conditions of the lattice. These are therefore determined from the equation of motion of the lattice, specialized for long wave vibrations (Born 1923). This leads to a secular equation for the velocity of propagation; from the diagonal terms of this equation one can obtain the elastic constants c_{11} and c_{44} in the case of cubic crystals. The off diagonal terms yield the sum $c_{12} + c_{44}$. Considering first the Coulomb part of the elastic constants, it is easy to show that $c_{44}^{(1)} = -\frac{1}{2} c_{11}^{(1)}$ where the superscript (1) is used to denote this part. The calculations for $c_{11}^{(1)}$ and $c_{44}^{(1)}$ were carried out by standard procedures, and the results are the same as those which have been obtained before (Born 1920, Kellermann 1940). The value of $c_{12}^{(1)} + c_{44}^{(1)}$ was also obtained independently. This serves as a check on the c_{12} value, which can also be obtained from the values of c_{11} and c_{44}, the form of the equations of motion, and from the Cauchy relations in the equilibrium state.

The values are

$$c_{11}^{(1)} = -1{\cdot}28 \qquad c_{44}^{(1)} = 0{\cdot}64 \qquad c_{12}^{(1)} = 0{\cdot}0560 \qquad \ldots\ldots (11)$$

the unit being e^2/r^4, where in this case r is half the cube spacing. Except for an inversion of $c_{12}^{(1)}$ and $c_{44}^{(1)}$ the first two values are those given by Kellermann (1940). The contributions from the repulsive potential were considered in two cases, in both of which the nth power law was used. In the first, nearest neighbours only were considered as this provides a particularly simple model. Here

$$c_{11}^{(2)} = \alpha(n+1)\, r_0^{n-1}/6r^{n+3} \qquad c_{44}^{(2)} = -\alpha r_0^{n-1}/6r^{n+3}$$
$$c_{12}^{(2)} = \alpha r_0^{n-1}/6r^{n+3} \qquad \ldots\ldots (12)$$

where α is the Madelung constant (1·747) and r_0 is the equilibrium value of r. In the more general case these have been considered in the region near isotropy, i.e. where n lies between 8 and 9.

In the isotropic case ($n=8\cdot55$) when all neighbours are taken into account the values are

$$c_{11}^{(2)}=2\cdot62\, r_0^{n-1}/r^{n+3} \qquad c_{44}^{(2)}=-0\cdot220\, r_0^{n-1}/r^{n+3}$$
$$c_{12}^{(2)}=0\cdot380\, r_0^{n-1}/r_0^{n+3}. \qquad \dots\ ..(13)$$

These two cases do give appreciably different results for γ_0 when n is chosen the same, and the question of using nearest neighbour forces is therefore of some significance.

§5.

The calculation of γ_0 in the isotropic case involves a certain amount of care as the solutions of the secular equation for the velocity become degenerate. The general equation can be written in the form (Born 1923)

$$z^2-z^3-z(1-K^2)S+(1-3K^2+2K^3)P=0 \qquad \dots\dots(14)$$

where $z=(\rho c^2-c_{44})/(c_{11}-c_{44}),\quad K=(c_{12}+c_{44})/(c_{11}-c_{44})$

$$S=(s_x^2s_y^2+s_y^2s_z^2+s_z^2s_x^2),\quad P=s_x^2s_y^2s_z^2.$$

Here c is the velocity of an elastic wave, ρ is the density of the crystal; c_{11}, c_{12}, c_{44} are the three elastic constants of the cubic crystal, and s_x, s_y, s_z are the components of the (unit) wave vector $\mathbf{s}$ associated with the plane wave. In the isotropic case $K=1$, leading to the solutions $z=1$ and $z=0$ (twice). The Cardan solution for the cubic equation is therefore expanded to the first order assuming $K-1$ to be small. The solutions then take the form

$$z_1=(1-K)[S+(S^2-3P)^{1/2}]$$
$$z_2=(1-K)[S-(S^2-3P)^{1/2}]$$
$$z_3=1-2K(1-S). \qquad \dots\dots(15)$$

It follows that

$$\frac{dz_1}{dr}+\frac{dz_2}{dr}=-2S\frac{dK}{dr}$$
$$\frac{dz_3}{dr}=2S\frac{dK}{dr} \qquad \dots\dots(16)$$

where r is here used as in §2.

Considering a particular wave vector $\mathbf{s}$, S and P will be unaltered as the volume is changed. For the two transverse waves (c_1 and c_2) and the longitudinal wave (c_3) associated with a particular $\mathbf{s}$ in the limit $K=1$, it follows (see (14)) that

$$\frac{dc_1^2}{dr}+\frac{dc_2^2}{dr}=-\frac{c_{11}-c_{44}}{\rho}2S\frac{dK}{dr}+\frac{d}{dr}\left(\frac{2c_{44}}{\rho}\right)$$
$$\frac{dc_3^2}{dr}=\frac{d}{dr}\left(\frac{c_{11}}{\rho}\right)+\frac{c_{11}-c_{44}}{\rho}2S\frac{dK}{dr}. \qquad \dots\dots(17)$$

The two transverse waves can be grouped together as their velocities will be equal in the limiting case; and the weighting factor in (9) is therefore the same. In fact $c_1^2=c_2^2=c_{44}/\rho$ and $c_3^2=c_{11}/\rho$. As $\int S d\Omega/4\pi=\frac{1}{5}$, it follows that

$$\int\left(\frac{dc_1^2}{dr}+\frac{dc_2^2}{dr}\right)\frac{d\Omega}{4\pi}=\frac{2}{5}\frac{d}{dr}\left(\frac{c_{11}-c_{12}+3c_{44}}{\rho}\right)=A$$
$$\int\frac{dc_3^2}{dr}\frac{d\Omega}{4\pi}=\frac{1}{5}\frac{d}{dr}\left(\frac{3c_{11}+2c_{12}+4c_{44}}{\rho}\right)=B \qquad \dots\dots(18)$$

and

$$\gamma_0=\frac{1}{3}+\frac{Ac_1^2+Bc_3^2}{2c_1^2+c_3^2}. \qquad \dots\dots(19)$$

The two isotropic cases considered above lead to $\gamma_0 = 1{\cdot}15$ $(n=7)$, $\gamma_0 = 1{\cdot}46$ $(n=8{\cdot}55)$, the first being the nearest neighbour approximation.

§6.

The high temperature value of $\gamma(\gamma_\infty)$ cannot be obtained directly without a very detailed knowledge of the variation of γ_ν over the spectrum. A method of avoiding this has been used by Barron (1955) who extrapolates to the quantity required from a knowledge of the quantities $\gamma(p)$ where

$$\gamma(p) = \frac{\Sigma\ \gamma_\nu \nu^p}{\Sigma_\nu \nu^p} = -\frac{1}{p}\frac{d \log \mu_p}{d \log V} \qquad \ldots\ldots(20)$$

where $\mu(p) = \dfrac{1}{3N}\sum_\nu \nu^p$.

These expressions can be obtained, for $p=2$, 4, etc., from the secular equation for the frequencies and are known in the case where the masses of the ions are equal. The value wanted is $\gamma(0)$ while $\gamma(-3)$ is equal to γ_0. It turned out that the extrapolation in this case is by no means clear cut as $\gamma(4) < \gamma(2)$ and $\gamma(-3) < \gamma(2)$. It appears therefore that the best assumption that can be made without a good deal more investigation is that $\gamma_\infty = \gamma(0) = \gamma(2)$. This would lead to

$$\gamma_\infty = 1{\cdot}500 \quad (n=7) \text{ case (i)}$$
$$\gamma_\infty = 1{\cdot}76 \quad (n=8{\cdot}55) \text{ case (ii)}.$$

The conclusion which one can draw by comparison with the low temperature values is that an appreciable drop in the γ value is to be expected when the temperature is reduced. The safest case to consider is the second $(n=8{\cdot}55)$ where this change is about 20%. It is not to be expected that the assumption of equal masses will be of importance in this connection. This leaves $\gamma(2)$ unchanged, and the effect on $\gamma(4)$ will not be considerable except in extreme cases. The calculations should therefore be a reasonable approximation in the case of rocksalt where the condition of isotropy is a not unfair first approximation. Where comparison with experiment is however attempted, it should be noted that the γ_∞ value (of 1·76) is an extreme high temperature limit. What would be expected is something less than this at room temperature with a slow drop towards a low temperature value 15% to 20% smaller. The only experimental γ value the author has been able to find is a high temperature value quoted by Born (1923) which is 1·59. This is in remarkably good agreement with what is to be expected, in view of the number of assumptions involved in the comparison of theory and experiment.

§7.

One can obtain some insight into the variation of the individual $\gamma_0^{(C)}$ (equation (9)) contributed by different types of waves, by considering wave vectors in selected directions in space. These are chosen to be along the cube axes, the face and space diagonals. Here the equation for the velocities is separable. The values are listed below, the repulsive part of the elastic constants being evaluated for nearest neighbours only.

There are two features in these results which are striking. One is the very large variations in the $\gamma_j^{(C)}$ values. The other is the occurrence of negative values associated with the shear modulus c_{44}. The weighted average value would in these cases certainly be positive, but the existence of this variation, which will presumably also arise when higher frequencies are considered, emphasizes the need for the definition of a suitable average value such as $\overline{\gamma_\nu}$, which shows that the quantity needed in the theory can be something which varies slowly with frequency even if the individual γ_ν values fluctuate in a disturbing manner.

Direction	Velocities	$\gamma_j^{(C)}(n=10)$	$\gamma_j^{(C)}(n=7)$
[100]	$\rho c_1^2=c_{44}$	$-0{\cdot}23$	$-0{\cdot}74$
	$\rho c_2^2=c_{44}$	$-0{\cdot}23$	$-0{\cdot}74$
	$\rho c_3^2=c_{11}$	2·90	2·73
[110]	$\rho c_1^2=c_{11}$	2·90	2·73
	$\rho c_2^2=\frac{1}{2}(c_{11}-c_{22})$	3·80	3·42
	$\rho c_3^2=\frac{1}{2}(c_{11}+c_{12}+2c_{44})$	1·65	1·33
[111]	$\rho c_1^2=\frac{1}{3}(c_{11}-c_{12}+c_{44})$	2·15	1·52
	$\rho c_2^2=\frac{1}{3}(c_{11}-c_{12}+c_{44})$	2·15	1·52
	$\rho c_3^2=\frac{1}{3}(c_{11}+2c_{12}+2c_{44})$.	1·62	1·32

REFERENCES

Barron, T. H. K., 1955, *Phil. Mag.* **46**, 720; 1957, *Annals Phys.*, in the press.

Born, M., 1920, *Ann. Phys., Lpz.*, **61**, 87; 1923, *Atomtheorie des festen Zustandes*, (Leipzig and Berlin: Teubner).

Kellermann, E. W., 1940, *Phil. Trans. Roy. Soc.* A, **238**, 513.

The Effect of the Equatorial Electrojet on the Ionospheric E_s and F2 Layers

By N. J. SKINNER and R. W. WRIGHT
Physics Department, University College, Ibadan, Nigeria

MS. received 2*nd April* 1957, *and in final form* 22*nd May* 1957

Abstract. It is found at Ibadan that during the daylight hours fE_s is lower on magnetically disturbed days than on quiet days. It is also found that on two or three occasions in each month a series of abnormally low values of fE_s occurs. An investigation of the behaviour of the horizontal and vertical components of the earth's field during these times reveals that there is a simultaneous decrease in the magnitude of the equatorial electrojet current which is in turn responsible for the large diurnal ranges in H and Z near the equator. The behaviour of the F2 layer parameters f_0F2, y_mF2 and h_mF2 during periods of low fE_s, shows that the F2 layer always becomes denser, lower and thinner during these periods.

This indicates an immediate connection between the electrojet, the production of equatorial E_s and vertical drifts in the F2 layer.

§ 1. INTRODUCTION

INFORMATION regarding the equatorial sporadic E layer E_s is particularly sparse and any correlations between the presence of E_s and any other geophysical features are therefore of interest.

The sporadic E layer near the equator is normally present during the daylight hours and the 'top frequency' fE_s follows a smooth diurnal variation with a maximum at noon. At any one time the layer virtual height is independent of frequency, although it varies diurnally with a minimum at noon (Skinner and Wright, in preparation). The layer is normally partially reflecting and appears as a diffuse trace on the (h', f) record, typical examples of which are reproduced in figure 1 (Plate). Blanketing only occurs when a markedly different type of E_s appears, usually at a much higher altitude. Even when 'blanketing E_s' is present, the transparent 'equatorial' type is usually still evident at a lower height of about 100 km. Since equatorial E_s is formed at the bottom of the normal E layer, distinction between h'E and $h'E_s$ is extremely difficult especially on the normal ionograms.

Matsushita (1951) has suggested that the presence of intense E_s in a narrow equatorial belt is linked with the large diurnal variation in the horizontal component of the earth's magnetic field H, which is in turn dependent upon the intense E–W S_q current (the so-called electrojet) known to flow at heights appropriate to the E layer near the equator (Egedal 1948, Singer, Maple and Bowen 1952). It is therefore to be expected that during periods of magnetic disturbance, changes should take place in the E_s layer. In §2 of this paper, a dependance of fE_s upon world-wide magnetic disturbance is established and it is found that fE_s is markedly reduced when disturbed magnetic conditions prevail.

The use of international magnetic character figures is rather unsatisfactory, however, since local variations, which may be large especially near the magnetic equator, tend to be smoothed out. Since November, 1955, a magnetic observatory has been operated by the Physics Department of University College, Ibadan, under the direction of Professor N. S. Alexander. Direct comparisons between the actual variations in the earth's horizontal and vertical field components, H and Z, at Ibadan and the simultaneous measurements of $f\mathrm{E}_s$ can thus be made. A study of the routine ionograms at Ibadan has revealed that for certain periods of a few hours during daylight, the E_s layer almost disappears. Such periods occur only about three or four times per month on an average, and indicate large local magnetic disturbance. A special investigation of the behaviour of H and Z at Ibadan during these periods has been made and the results are reported in § 3.

It is well established that electrodynamical forces giving rise to vertical electron drifts are associated with the electric currents flowing in the ionosphere. These drifts have been shown (Skinner and Wright 1955) to perturb the F2 layer at Ibadan. In § 4, the behaviour of the F2 region parameters f_0F2, h_mF2 and y_mF2 is examined for periods of abnormally low values of $f\mathrm{E}_s$.

§ 2. $f\mathrm{E}_s$ ON MAGNETICALLY QUIET AND DISTURBED DAYS

Matsushita 1951 has suggested that there is a connection between the large diurnal range in H near the magnetic equator and the presence of intense sporadic E. On magnetically disturbed days the diurnal range in H is usually reduced at Ibadan so that one would expect less E_s on these days. In order to verify this experimentally, the mean diurnal variation of $f\mathrm{E}_s$ is plotted for the internationally selected magnetically quiet days (curve Q) and the similarly selected disturbed days (curve D) in figure 2. These curves are based on one year's data, August 1955–July 1956. It is seen that during the daylight hours 0900–1500 GMT, $f\mathrm{E}_s$ is more than 1 Mc/s larger on quiet days than on disturbed days.

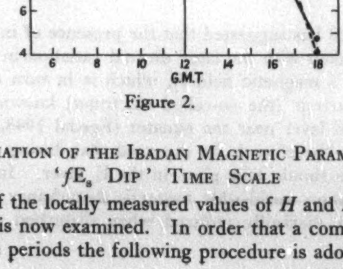


Figure 2.

§ 3. THE VARIATION OF THE IBADAN MAGNETIC PARAMETERS ON AN '$f\mathrm{E}_s$ DIP' TIME SCALE

The behaviour of the locally measured values of H and Z during periods of abnormally low $f\mathrm{E}_s$ is now examined. In order that a common time scale can be used for all these periods the following procedure is adopted. The time of commencement of the pronounced dip in fE_s is selected as the zero of a time scale which we shall call the 'fE_s dip' time scale. The period of six months November 1955 to April 1956 contains 20 such occurrences. The magnetic parameters have been examined during the interval from four hours before to seven hours after the zero time on these 20 occasions. For each period the differences between the hourly values of H and Z and the monthly mean values for those hours, $H_t-\overline{H}_t$ and $Z_t-\overline{Z}_t$, were calculated and arranged in 'fE_s dip' time. The mean dip time variation of these quantities is shown in figure 3(*a*) and (*b*) together with the dip time variation of $fE_s/\overline{fE_s}$ (figure 3(*c*)).

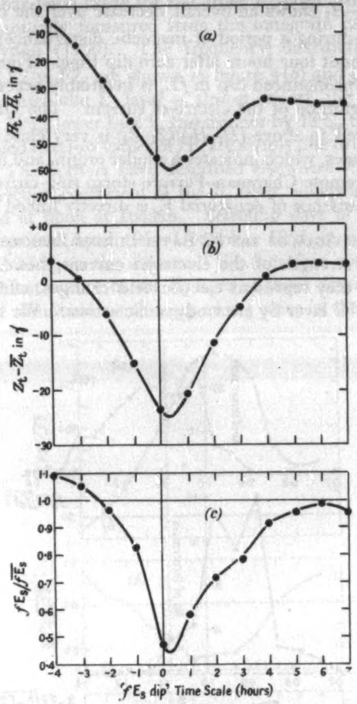


Figure 3.

A remarkable similarity between the curves in figure 3 (*a*), (*b*) and (*c*) is at once apparent. A sudden decrease in fE_s is invariably accompanied by a similar sharp decrease in the magnetic field at the earth's surface. Each individual occurrence also shows this effect.

The diurnal ranges of both H and Z are abnormally large at Ibadan, the H variation being very similar to that observed at Huancayo (McNish 1937). Egedal (1948) has shown that such large diurnal ranges are restricted to a narrow belt of latitudes about 300 km wide and centred on the magnetic equator. He explains the effect in terms of an intense electric current flowing East–West at the equator, probably at E-layer heights. Using a rocket-borne magnetometer at the magnetic equator, Singer, Maple and Bowen (1952) have confirmed that a region of greatly enhanced conductivity exists at an altitude of about 100 km. At Ibadan, which is two or three degrees south of the magnetic equator, the diurnal ranges of H are smaller than those of Huancayo, but those of Z are much larger. The electrojet is thus not directly overhead at Ibadan and the changes in Z represent changes in the actual strength of the electrojet current. This is confirmed to some extent by a consideration of figure 3(*a*) and (*b*) where it is seen that whilst $H_t-\overline{H}_t$ shows an overall decrease over the complete period (as would be expected during a period of magnetic disturbance), $Z_t-\overline{Z}_t$ recovers to its normal values about four hours after zero dip time. Figure 3(*a*), (*b*) and (*c*) thus show that the pronounced dip in $f\text{E}_s$ is invariably accompanied by a large decrease in the magnitude of the electrojet current The ratio of the magnetic field changes referred to above $(H_t-\overline{H}_t)/(Z_t-\overline{Z}_t)$ is very close to the ratio of the normal diurnal changes, which indicates a similar origin, and no likely connection to the much more remote Chapman–Ferraro storm ring-current. We conclude therefore that the presence of equatorial E_s is directly linked to the electrojet.

§ 4. THE BEHAVIOUR OF THE F2 LAYER DURING PERIODS OF LOW $f\text{E}_s$

Changes in the strength of the electrojet current, besides affecting the E_s layer, should, since they represent electric field changes, affect the distribution of ionization in the F2 layer by electrodynamic action. We have examined this effect in two ways.

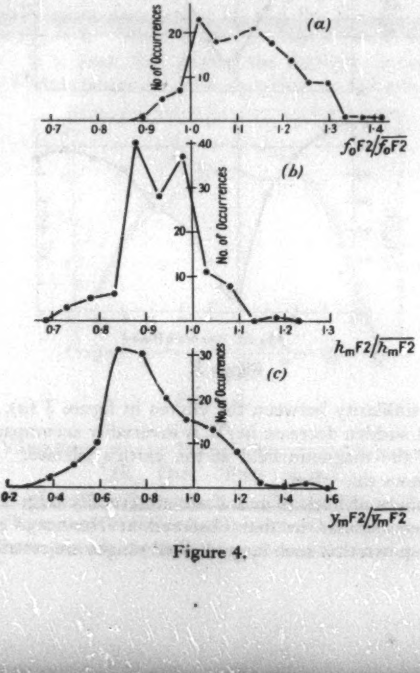


Figure 4.

4.1. *The Distribution of the F2 Layer Parameters during Periods of Low fE_s*

Data for the year May 1955–April 1956 have been studied. The periods of low fE_s were selected by taking hours on which fE_s was less than 60% of the monthly mean value for those times of day, together with the two or three adjacent hours for which fE_s was still considerably smaller than the monthly means. The ratio of f_0F2 for each of these hours to the corresponding monthly mean value for each particular hour, $\overline{f_0F2}$, has been calculated. The frequency distribution of this ratio $f_0F2/\overline{f_0F2}$ is plotted in figure 4(*a*). It is seen that, on the average, f_0F2 is increased by about 12% on hours for which fE_s is abnormally small. A similar analysis has been made for the other F2-layer parameters, h_mF2 and y_mF2, both being initially measured from the ionogram using the method of Appleton and Beynon (1940). The frequency distributions of the ratios, $h_mF2/\overline{h_mF2}$ and $y_mF2/\overline{y_mF2}$, are shown in figure 4(*b*) and (*c*). It is seen that, during periods in which the E_s layer is almost absent, the F2 layer is, on the average, denser by 24%, lower by 6% and thinner by 18%, than normal.

It might be argued that low values of fE_s are simply due to increased absorption in the D-region. The authors have described elsewhere (Skinner and Wright 1957) how the reflection coefficient of the E_s layer, independent of absorption, has been measured at noon at Ibadan. Selecting days of very low reflection coefficient for the period April 1954–December 1955, it is found that f_0F2 is increased by about 13% on the average. We therefore conclude that there is a real change in the structure of the E_s layer at these times which is linked in some way with the simultaneous changes taking place in the F2 layer.

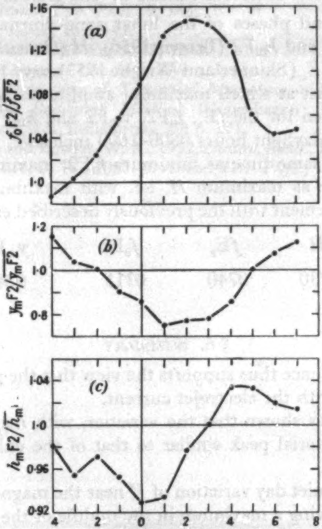


Figure 5.

An alternative approach to the same study has also been made. The correlation coefficient between the values of f_0F2 and fE$_s$ for the three hours 1100, 1200 and 1300 GMT has been calculated for a period of two months. A coefficient of $-0{\cdot}35$ has been obtained based on 160 readings, a result which is significant to better than the 1% level. The method first described has the merit that the periods with exceptionally small values of fE$_s$ can be selected from the ionograms, ensuring that the low value of fE$_s$ is not due to an instrumental fault or to anomalously large absorption.

4.2. *The Dip Time Variation of the F2 Layer Parameters*

Values of the ratios $f_0\mathrm{F2}/\overline{f_0\mathrm{F2}}$, $y_m\mathrm{F2}/\overline{y_m\mathrm{F2}}$ and $h_m\mathrm{F2}/\overline{h_m\mathrm{F2}}$ have been arranged in 'fE$_s$ dip' time and the mean dip time variations are plotted in figure 5 (*a*), (*b*) and (*c*). The time of maximum effect in the F2 layer is seen to be delayed by about 2 hours, with an increase in f_0F2 of about 14% and a decrease in thickness of 25%.

The associated changes in h_mF2 are more complex showing an initial decrease of 7% followed by an increase to about 4% above average. This initial decrease is in agreement with that observed in figure 4 (*b*) which was based upon data for the three or four hours centred about zero 'fE$_s$-dip' time. It is possible that the increase in h_mF2 commencing 2 hours after zero dip time may not be real since the measurement of this parameter by the Appleton–Beynon method is particularly affected by distortion of the layer.

§ 5. EVIDENCE FROM THE RELATIVE PHASES OF THE LUNAR SEMIDIURNAL TIDES

The amplitudes and phases of the lunar semi-diurnal tidal oscillations at Ibadan in f_0F2, y_mF2 and h_mF2 (Brown 1956), *H* (Onwumechilli 1957, private communication) and fE$_s$ (Skinner and Wright 1957) have been calculated. The times after lunar transit at which maximum amplitudes occur are given in the table. The times given for the fE$_s$, f_0F2, y_mF2 and h_mF2 tides relate to data measured during the daylight hours 0800–1600 inclusive. Maximum values of fE$_s$ thus occur at the same time as minimum f_0F2, maximum y_mF2 and h_mF2, and at the same time as maximum *H*, i.e. with maximum electrojet current. This is in general agreement with the previously described experimental evidence.

Parameter	H	fE$_s$	f_0F2	y_mF2	h_mF2
Time of max. amplitude	0730	0740	0315	0840	0745

§ 6. SUMMARY

The following evidence thus supports the view that the presence of equatorial E$_s$ is directly linked with the electrojet current.

(i) Matsushita has shown that the variation with magnetic latitude of fE$_s$ shows a marked equatorial peak similar to that of the enhanced diurnal range of *H*.

(ii) The normal quiet day variation of *H* near the magnetic equator is similar to that of fE$_s$, both having a maximum in the middle of the day.

(iii) During periods of abnormally low values of fE$_s$ at Ibadan, there is a simultaneous fall in *H* and *Z* indicating a drop in electrojet current strength.

(iv) Further evidence that the electrojet current declines during these periods is demonstrated by the behaviour of the F2 layer at these times.

(v) The relative phases of the lunar semi-diurnal oscillations in f_0F2, y_mF2, h_mF2, H and fE$_s$ at Ibadan are in agreement with the main conclusions above.

The exact mechanism by which equatorial E$_s$ reflections occur is still not established. It is possible that electrodynamical forces set up in the E layer by the flow of the electrojet current may cause the layer to have a sharp base with a very high vertical gradient of ionization density. Any turbulence present at this sharp boundary might set up major scattering centres sufficient to give rise to the observed equatorial E$_s$ reflection. It is interesting that Gallet (1955) has also considered this type of mechanism as a possible source of E$_s$ stratifications near the bottom of the E layer.

Alternatively the flow of the electrojet current itself may set up additional turbulence in the base of the E region.

However, whatever mechanism is responsible for the formation of the equatorial E$_s$ layer, the experimental evidence is overwhelmingly in favour of the view that it must be intimately connected with the presence of the electrojet current.

ACKNOWLEDGMENT

We wish to acknowledge many useful discussions with Dr. A. J. Lyon.

REFERENCES

APPLETON, E. V., and BEYNON, W. J. G., 1940, *Proc. Phys. Soc.*, **52,** 518.
BROWN, R. A., 1956, *J. Atmos. Terr. Phys.*, **9,** 144.
EGEDAL, J., 1948, *Nature, Lond.*, **161,** 443.
GALLET, R., 1955, *Proc. Inst. Radio Engrs*, **43,** 1240.
MATSUSHITA, S., 1951, *J. Geomagn. Geoelect., Kyoto*, **3,** 44.
MCNISH, A. G., 1937, *Association of Terrestrial Magnetism and Atmospheric Electricity* (*Transactions of the Edinburgh Meeting*) (Copenhagen: International Union of Geodesy and Geophysics, Bulletin no. 10) p. 271.
SINGER, S. F., MAPLE, E., and BOWEN, W. A., 1952, *Nature, Lond.*, **170,** 1093.
SKINNER, N. J., and WRIGHT, R. W., 1955, *J. Atmos. Terr. Phys.*, **6,** 177.

Time Variations of Extensive Air Showers

By F. J. M. Farley† and J. R. Storey‡
Auckland University College, Auckland, New Zealand

MS. received 9th May 1957

Abstract. Extensive air showers corresponding to primary cosmic rays ranging from 10^{14} to 3×10^{15} ev were recorded at Auckland for the year 1955. The results do not confirm the large sidereal diurnal effect reported previously.

§ 1. Introduction

In a previous publication (Farley and Storey 1954) we reported a sidereal diurnal amplitude of $(1{\cdot}10 \pm 0{\cdot}26)\%$ for extensive air showers of energy about 10^{14} ev observed at Auckland (altitude 40 m) in 1951. Further more accurate observations, carried out in 1955 and now presented, do not confirm this effect.

Since our first paper several laboratories have reported observations on the sidereal correlation of extensive air showers, but the effects, if any, are all of rather low statistical certainty (Cranshaw and Galbraith 1954, Citron and Kehler 1955, Auger, Cachon, Daudin, Freon and Moszkowski 1955, McCusker 1955, Crawshaw and Elliot 1956). Results up to 1954 have been reviewed by Greisen (1956). The very variability of the results obtained by different workers throws doubt on the existence of a genuine sidereal correlation.

In our new experiment improved statistics are obtained by using a larger apparatus, and showers of four different energies are selected for recording. Taking the original shower energy as E (estimated as 10^{14} ev), originally recorded at 50 counts per hour, we now record showers of energy E at 120/h, showers of energy $3E$ at 80/h, showers of energy $9E$ at 14/h, and showers of energy $30E$ at 3/h.

§ 2. Apparatus

Nine trays of Geiger counters were used as shown in the figure, each tray containing 10 counters, Cintel type GM 41A. In these counters ethyl formate is used as quenching vapour, giving better temperature stability than the alcohol-quenched counters used previously. The counters were also quenched externally in pairs by quenching circuits of improved design.

As shown in the figure, the trays were grouped in threes at the corners of an equilateral triangle of 51 m side, the three trays in each group being spaced 4·26 m between centres. From this array the following coincidences were recorded:

A. Three-fold coincidences 123, 456 and 789 are recorded separately and then combined. Because of the spacing of the groups these local coincidences at each corner of the main triangle are largely independent so that the combined counting rate is nearly three times that of each group separately. The shower energy selected is close to that of the original experiment in which similar trays were used at the corners of a 4 m equilateral triangle.

---

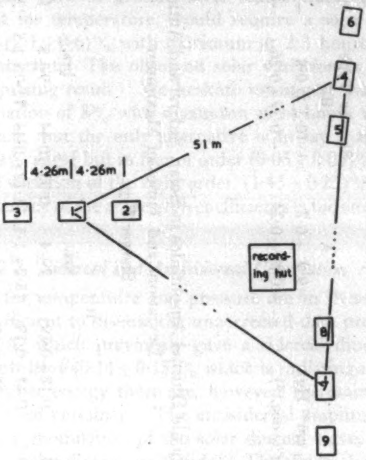
† Now at CERN, Geneva. ‡ Now at University of Tasmania, Hobart.

B. Three-fold coincidences (1+2+3), (4+5+6), (7+8+9) over the large triangle. The three trays at each corner were combined to form in effect a single tray of three times the area by adding the signals into a single channel. B-coincidences were then taken between the three channels so formed. According to calculations using the Molière structure function the shower energy is approximately $3E$.

C. Three-fold coincidences 147, using only one selected tray at each corner of the main triangle. The shower energy is therefore three times that for the B-coincidences, that is $9E$.

D. Nine-fold coincidences between all trays. The energy in this case is estimated from the counting rate and the primary spectrum. Assuming that the apparatus is sensitive to primary cosmic rays striking the same target area as for C-coincidences, the energy is about $30E$.

As before the data were recorded on a 20-pen operation recorder, and the counting rates of the individual trays were monitored throughout the experiment.



§ 3. ANALYSIS AND RESULTS

Results for one year beginning 10th December 1954 were selected from the recorded data and analysed for diurnal variations in solar, sidereal and 'antisidereal' time. Barometric and temperature coefficients were obtained by means of a double regression analysis (Hoel 1946) with the intention of correcting the diurnal coefficients for the effects of the mean diurnal waves of atmospheric pressure and temperature. For this purpose, hourly readings of sea level temperature and pressure were obtained from a meteorological station 1 mile away.

In finding these regression coefficients it is important to avoid errors due to the possible presence of a real diurnal variation of the cosmic rays not due to atmospheric influences. For example, if the cosmic ray count and the temperature

Table 1. Regression Coefficients Calculated for the Various Events Observed with respect to Barometric Pressure and Temperature at Sea Level

	Shower Type			
	A	B	C	D
Barometric coefficient (% cm⁻¹ Hg)	−9·9 ±0·7	−10·5 ±0·8	−14·2 ±1·8	−13·6±3·7
Temperature coefficient (% deg.⁻¹)	−0·89±0·24	−0·37±0·26	−0·37±0·60	−2·1±1·3

Table 2. First Harmonic Diurnal Amplitudes in Solar, Sideral and Antisidereal Time with Standard Deviations and Times of Maximum

	Shower Type								Pressure		Temperature	
	A		B		C		D					
	%	hrs.	%	hrs.	%	hrs.	%	hrs.	(mb)	hrs.	(°c)	hrs.
Solar	0·13±0·15	19 ±4	0·18±0·19	2±4	0·33±0·44	9±5	1·1±1·0	15±3	0·24	6·1	2·42	14·3
Sidereal	0·14±0·15	14·3±4·1	0·47±0·19	15·2±1·5	1·12±0·44	12·0±1·5	2·3±1·0	10·8±1·6	0·00	—	0·25	18·0
Antisidereal	0·30±0·15	10±2	0·07±0·19	3±10	1·01±0·44	11±2	0·9±1·0	14±4	0·07	14·2	0·38	8·9

were both independently correlated with solar time a spurious temperature coefficient might be obtained. To avoid such an effect only the data recorded from 7.30 p.m. to 12.30 a.m. was used each day for the regression calculation, the analysis being confined to four months, spaced through the year. As a result the accuracy of the regression coefficients is limited, especially in the case of temperature because the temperature change from day to day was small. The values obtained are given in table 1.

The analysis for first harmonic diurnal amplitude was made following the procedure described in our first paper with the results given in table 2. These figures have not been corrected for pressure and temperature variation, but the corrected amplitudes can be obtained using the regression coefficients of table 1 together with the mean diurnal amplitudes of pressure and ground level temperature given in table 2.

3.1. *Solar Correlation*

The mean diurnal wave of ground level temperature, together with the regression coefficient for temperature, would require a solar diurnal variation of A-type events of $(2{\cdot}1 \pm 0{\cdot}6)\%$ with maximum at 2·3 hours, if the incoming cosmic rays were constant. The observed solar variation is however close to zero. This is a surprising result. We hesitate to suggest that the cosmic rays have a real solar variation of 2% with maximum at 14 hours which just cancels the atmospheric effect; but the only alternative is to say that the temperature coefficient is not $0{\cdot}89\%\ \mathrm{deg}^{-1}$ but in fact of order $(0{\cdot}05 \pm 0{\cdot}05)\%\ \mathrm{deg}^{-1}$. In 1951 we did obtain a solar variation of the right order, $(1{\cdot}45 \pm 0{\cdot}25)\%$. For the higher energy events the accuracy of the regression coefficients is too small for conclusions to be drawn.

3.2. *Sidereal and Antisidereal Correlation*

The corrections for temperature and pressure are in these cases practically negligible, so it is sufficient to discuss the uncorrected data presented in table 2. For events of type A, which previously gave a sidereal diurnal amplitude of $(1{\cdot}10 \pm 0{\cdot}26)\%$ we now have $(0{\cdot}14 \pm 0{\cdot}15)\%$ which is indistinguishable from zero. For the events of higher energy there are, however, increasing amplitudes but all of very low statistical certainty. The antisidereal amplitudes, which are in theory due to seasonal modulation of the solar diurnal wave, should always be less than half the mean solar diurnal amplitude. The results, however, show large antisidereal amplitudes, which can only be ascribed to statistical fluctuations. It is therefore meaningless to attempt to apply corrections for the solar–seasonal modulation effect. The occurrence of these large antisidereal amplitudes also suggests strongly that the sidereal amplitudes themselves may well have arisen from a statistical fluctuation.

We conclude quite definitely that the large sidereal amplitude for events of type A obtained in 1951 was not present in 1954, while evidence for an effect at higher energies is very weak.

ACKNOWLEDGMENT

Funds for this experiment were provided by the University of New Zealand Research Committee. We wish to thank Professor P. W. Burbidge for his continued interest and encouragement.

REFERENCES

AUGER, P., CACHON, A., DAUDIN, A., FREON, A., and MOSZKOWSKI, C., 1955, *C.R. Acad. Sci., Paris*, **240**, 2407.

CITRON, A., and KEHLER, P., 1955, *Z. Naturf.* **10a**, 499.

CRANSHAW, T. E., and GALBRAITH, W., 1954, *Phil. Mag.*, **45**, 1109.

CRAWSHAW, J. K., and ELLIOT, H., 1956, *Proc. Phys. Soc.* A, **69**, 102.

FARLEY, F. J. M., and STOREY, J. R., 1954, *Proc. Phys. Soc.* A, **67**, 996.

GREISEN, K., 1956, *Progress in Cosmic Ray Physics*, Vol. III (Ed. J. G. Wilson) (Amsterdam : North-Holland).

HOEL, P. G., 1946, *Introduction to Mathematical Statistics* (London : Chapman and Hall). p. 110.

MCCUSKER, C.B.A., 1955, Nuovo Cim., **2**, 1340.

The Measurement of the Optical Properties of Zinc Sulphide

By C. K. COOGAN

Chemical Physics Section, Division of Industrial Chemistry, Commonwealth Scientific and Industrial Research Organization, Fishermen's Bend, Melbourne.

Communicated by A. L. G. Rees; MS. received 6th September 1956, *and in revised form* 1*st April* 1957

Abstract. The knowledge, to date, of optical absorption in ZnS is surveyed and discussed. The method used for determining the optical absorption and refractive index of a solid over a range of wavelengths is discussed and equations deduced for the transmittance and reflectance of a thin absorbing film on a thick substrate. The practical determination of the refractive index and extinction coefficient from the measured values of transmittance and reflectance is discussed in some detail. A description is given of the experimental procedure and precautions used in determining the optical constants of evaporated layers of ZnS. The results are discussed in relation to previous work. The effects of change of temperature at which the measurements are made, and of exposure of the films to air are noted and discussed.

---

§ 1. INTRODUCTION

ALTHOUGH there is considerable interest in the physics of the solid state, there is little quantitative information on the energy diagrams of ionic and semi-conducting solids, even for the simplest of such systems, the alkali halides. Such information must come almost exclusively from measurement of the optical constants, including the intrinsic absorption. Many of the models used for the interpretation of the electronic properties of solids, particularly optical and luminescence properties and electronic conduction, have been based on speculation as to the energy diagram. It is evident that there is a wide need for this information and for rapid and precise methods of measurement of the parameters from which the energy diagram may be deduced.

Zinc sulphide has been chosen for these investigations, since it is representative of a class of solids showing a good deal of covalence in their electronic binding and many of its electronic properties have been studied extensively.

Despite the interest of the solid state physicist in zinc sulphide, the energy level diagram of pure zinc sulphide has never been established.

Seitz (1940) gives 10·5 ev as a theoretical estimate of the separation of the valence and conduction bands, whereas Wright (1948) gives 5·2 ev; estimates based on experimental measurement assume that the long wavelength edge of the strong absorption in the near ultra-violet region corresponds to the transition of lowest energy between the valence and conduction bands. A current theory of quenching processes in the luminescence of ZnS phosphors requires that the energy gap (valence band to conduction band) is of the order of 3·5 ev, whereas other mechanisms are possible with much larger values for this energy gap. This divergence of estimated values suggests that an unambiguous determination of this energy in particular and the whole diagram in general would be invaluable.

Considerable progress may be made towards a solution of this problem by measuring the absolute absorption coefficient throughout the strongly absorbing region, to as high energies as practicable, and at the same time to measure, as a function of the energy, another property (e.g. refractive index, photoconductivity at low temperature) which serves to distinguish transitions to bound states from transitions to free states. In doing this, experimental problems attributable directly to the high absorption coefficients ($\sim 10^6$ cm^{-1}) arise.

This paper describes the measurement and computation of the relevant optical constants of ZnS at 20°c and -186°c in the spectral range 230–700 mμ, for which novel methods have been devised. The interpretation of results and the implications thereof in the theory of luminescence phenomena will be discussed in another paper.

Some previous work has been carried out on the absorption of ZnS. Gisolf (1939) has measured the relative absorption of some ZnS phosphors, using thin powder layers. While useful in revealing the structure of activator defect absorption bands qualitatively in the long wavelength tail region, this method is unreliable. Brasefield (1940) also made some measurements on single crystals, in the long wavelength region.

Because of high absorption in the ultra-violet, measurements must be made on very thin layers, which can only reasonably be obtained in polycrystalline films, prepared by evaporation at low pressure, sublimation or chemical deposition on to transparent substrates. Measurements on thin films are complicated by the fact that optical interference takes place between the front surface and the film–substrate interface. When the transmittance of the film is less than 0·1 the error introduced by neglecting the effects of interference will be negligible compared with the experimental error. However, for transmittances greater than 0·1 interference must be taken into account. Mollwo (1944) has measured the absolute absorption constant of ZnS using a thin evaporated layer, neglecting interference so that his results will be in error near the absorption edge. Similarly, Shalimova (1951) has investigated the relative absorption of sublimated ZnS phosphors, containing various activators, at room temperature. Kuwabara and Isiguro (1952) have obtained both the refractive index and extinction constant of a number of evaporated layers of ZnS at room temperature, from 900 to 400 mμ, taking the interference into account.

More recently, Hall and Ferguson (1955a,b) have measured the refractive index and absolute absorption constant of evaporated layers of ZnS for wavelengths down to 400 mμ.

We have developed apparatus and a technique for obtaining the refractive index and the extinction constant of ZnS at room temperature, or at liquid air temperature. The technique is applicable to other materials and offers scope for further work in obtaining similar much needed data for the silver halides, and other important classes of solids. The apparatus, technique and results are described in the following sections.

§ 2. Interference and Absorption in Thin Films

The value of the absorption constant for most solids lies in the range 10^5–10^6cm^{-1} for absorption processes intrinsic to the crystal matrix and not due to impurities or structural defects, etc. Thus the transmissivity T for two particular film thicknesses is given below:

Absorption constant (cm^{-1})	10^5	10^6	10^5	10^6
Film thickness (Å)	1000	1000	5000	5000
Transmissivity T	$10^{-0{\cdot}43}$	$10^{-4{\cdot}34}$	$10^{-2{\cdot}17}$	$10^{-2{\cdot}17}$

The necessity for using thin films is thus clearly seen. When thin films are used interference is important and its effects have to be calculated. We will assume a model in which the film is composed of an isotropic homogeneous layer of uniform refractive index on a non-absorbing isotropic, homogeneous substrate. For a non-absorbing film of thickness h_2 and refractive index n_2, deposited on a thick substrate of refractive index n_3, the expressions for the maximum and minimum reflectivity R at normal incidence in air, if $n_2 > n_3$, are

$$R_{\max} = (n_2{}^2 - n_3)^2/(n_2{}^2 + n_3)^2 \qquad \ldots\ldots(1)$$

$$R_{\min} = (n_3 - 1)^2/(n_3 + 1)^2. \qquad \ldots\ldots(2)$$

$R_{\max}$ and $T_{\min}$ occur when

$$n_2 h_2 = \tfrac{1}{4}\lambda(2m+1) \qquad \ldots\ldots(3)$$

and $R_{\min}$ and $T_{\max}$ occur when

$$n_2 h_2 = \tfrac{1}{4}\lambda\, 2m, \qquad \ldots\ldots(4)$$

where $m+1$ is the order of the interference and λ is the wavelength of the incident light *in vacuo*. $R_{\min}$ is equal to the reflection from the substrate alone for n_2 greater than n_3.

When there is absorption in the layer, the expressions for reflectivity and transmissivity take the form

$$R = F_{\mathrm{R}}(n_2, K_2, h_2, n_3, \lambda, \phi,) \qquad \ldots\ldots(5)$$

$$T = F_{\mathrm{T}}(n_2, K_2, h_2, n_3, \lambda, \phi,), \qquad \ldots\ldots(6)$$

assuming the layer to be composed of isotropic material, where K_2 is the absorption constant for the film, λ the wavelength of incident light *in vacuo* and ϕ the angle of incidence. In addition R and T depend on the state of polarization of the incident light. For small angles of incidence, the dependence on ϕ and the state of polarization disappears, to a good approximation, and we can use expression involving n_2, K_2, h_2, n_3 and λ only. Figure 1 shows the layer order. Mooney (1945) has derived expressions for R_0 and T_0, the normal incidence reflectivity and transmissivity respectively for a one-layer film without absorption, and for a two-layer film with and without absorption.

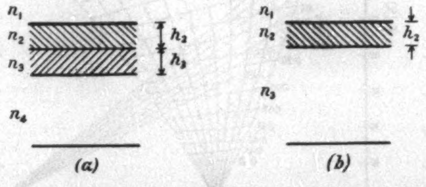


Figure 1. Order of the layers: (*a*) two-layer surface used in Mooney's calculations, (*b*) one-layer surface used here. In both cases $n_1 = 1$ (for air).

If we make the refractive indices of the two layers equal, the thickness and absorption constant of the lower layer zero, and the absorption constant of the upper layer not zero, we get the equations for one absorbing layer only, on a semi-infinite substrate, i.e.

$$R_0 = \frac{[(n_2-1)(n_2+n_3)]^2 \exp(k_2h_2)+[(n_2+1)(n_3-n_2)]^2 \exp(-k_2h_2)+2(n_2^2-1)(n_3^2-n_2^2)\cos y}{[(n_2+1)(n_2+n_3)]^2 \exp(k_2h_2)+[(n_2-1)(n_3-n_2)]^2 \exp(-k_2h_2)+2(n_2^2-1)(n_3^2-n_2^2)\cos y} \quad \ldots\ldots(7)$$

where K_2 is the absorption constant for the layer, n_2 the refractive index for the layer, h_2 the thickness of the layer, $y = 4\pi n_2h_2/\lambda$ and n_3 is the refractive index for the substrate.

$$T_0 = 16n_2^2\, n_3/Q' \quad \ldots\ldots(8)$$

where Q' is the denominator in the expression for R.

It is to be noted that the extinction constant k_2 of the complex refractive index $n_2 - ik_2$ is related to K_2 by the expression $k_2 = \lambda K_2/4\pi$.

Unfortunately, given R, T, h_2, n_3 and λ, we cannot derive explicit equations for n_2 and k_2 in terms of R_0, T_0, n_3, h_2 and λ owing to the algebraic complexity of the equations. We are forced to use some variation of 'cut-and-try' methods, as used by Brattain and Briggs (1949), Kuwabara (1954) and Hass and Saltzberg (1954), for example.

We can also see from the functional form of R_0 and T_0 in equations (5) and (6) that

$$\frac{dR}{d\lambda} = \left(\frac{\partial R}{\partial \lambda}\right) + \left(\frac{\partial R}{\partial n_2}\right)\frac{dn_2}{d\lambda} + \left(\frac{\partial R}{\partial n_3}\right)\frac{dn_3}{d\lambda} + \left(\frac{\partial R}{\partial K_2}\right)\frac{dK_2}{d\lambda}. \quad \ldots\ldots(9)$$

If n_2, n_3 and K_2 vary slowly with respect to change in λ then

$$\frac{dR}{d\lambda} \simeq \left(\frac{\partial R}{\partial \lambda}\right). \quad \ldots\ldots(10)$$

Thus $R_{\max}$ and $R_{\min}$ occur at almost the same positions that we had before in equations (3)and(4), except near absorption edges or near regions of anomalous

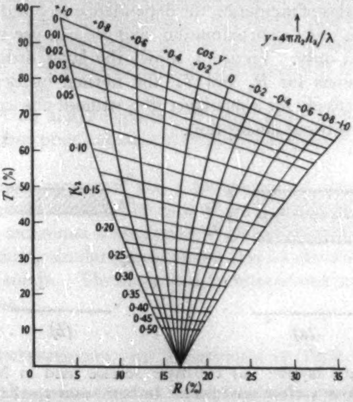


Figure 2. Typical member of the set of graphs showing T as a function of R for given combinations of n_2 and n_3 and varying values of absorption constant. Here $n_2 = 2{\cdot}4$, and $n_3 = 1{\cdot}46$.

dispersion where $\partial K_2/\partial\lambda$ and $\partial n_2/\partial\lambda$ are appreciable. This enables us to use the simple expressions linking refractive index and thickness, particularly with moderately thick specimens where the maxima and minima occur at smaller wavelength intervals, and hence refer to smaller changes in n_2, n_3 and K_2.

We have calculated R_0 and T_0 over a range of the parameters n_2, K_2, n_3 and $\cos y$ and plotted the results for easy reference. The refractive index for fused silica varies from 1·455 at 700 mμ to 1·470 at 400 mμ. We have chosen the value 1·460 as a sufficiently close approximation over this range, in which the ZnS absorption is small. Using $n_3 = 1{\cdot}46$, we have chosen a set of convenient values of n_2 covering the range $n_2 = 2{\cdot}2$–$2{\cdot}7$ in steps of 0·1. For each of these values of n_2, we have calculated the values of R_0 and T_0 for a set of values of K_2h_2 using $\cos y$, which ranges from $+1$ to -1, as the variable parameter. Then for each separate

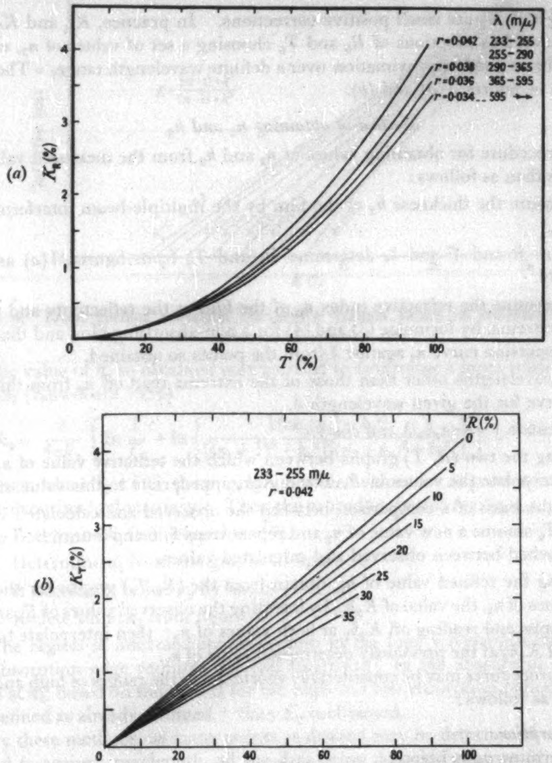


Figure 3. Curves for correcting apparent reflection and transmission for back surface reflection : (*a*) reflection correction, (*b*) transmission correction.

combination of n_2 and n_3, T_0 is plotted as a function of R_0 for constant K_2h_2 and constant cos y. Figure 2 shows a specimen of this set of graphs. As a corollary of equation (10), extrema occur where $\cos y = \pm 1$.

The measured values of R and T will differ from R_0 and T_0 owing to reflection from the back surface of the substrate. To a sufficiently good approximation,

$$R \simeq R_0 + T_0^2 r (1 + rR_0), \qquad \ldots\ldots(11)$$

$$T \simeq T_0 (1 - r)(1 + rR_0), \qquad \ldots\ldots(12)$$

where R and T are the measured reflectance and transmittance of the specimen and r is the reflection of the back surface. Since the dispersion curve for the fused silica is known, r can be calculated, and the above equations put in the form

$$R_0 = R - K_R \qquad \ldots\ldots(13) \qquad T_0 = T + K_T, \qquad \ldots\ldots(14)$$

where K_R and K_T are small positive corrections. In practice, K_R and K_T have been calculated as functions of R_0 and T_0 choosing a set of values of n_3, each of which will give good approximation over a definite wavelength range. These are plotted as in figures 3 (*a*) and (*b*).

Method of obtaining n_2 and k_2

The procedure for obtaining values of n_2 and k_2 from the measured values of R and T is then as follows:

1. Measure the thickness h_2 of the film by the multiple-beam interferometry technique.

2. From R and T and λ, determine R_0 and T_0 from figures 3 (*a*) and (*b*) respectively.

3. Determine the refractive index n_2 of the film at the reflectivity and transmissivity extrema, by formulae (3) and (4) for a non-absorbing film, and thus plot a rough dispersion curve n_2 against λ from the points so obtained.

4. At wavelengths other than those of the extrema read off n_2 from this dispersion curve for the given wavelength λ.

5. Calculate $y = 4\pi n_2 h_2/\lambda$ and cos y.

6. Using the two (R, T) graphs between which the tentative value of n_2 lies, linearly interpolate the values of R, T and K_2h_2 appropriate to this value of n_2.

7. On the basis of a comparison between the observed and calculated values of R_0 and T_0 assume a new value of n_2 and repeat steps 5, 6 and 7 until best agreement is reached between observed and calculated values.

8. Using the refined value of n_2, obtain from the (R_0,T_0) graphs for the two nearest values of n_2, the value of K_2h_2, by inserting the observed values of R_0 and T_0 in both graphs and reading off K_2h_2 at both values of n_2; then interpolate to find the value of K_2h_2 at the previously determined value of n_2.

These procedures may be considerably shortened in the ranges of high and low absorption as follows:

I. *Low absorption.*

1. Determine n_2 as before.

2. It is found from the graphs that K_2h_2 for a given R_0, T_0 varies only slightly with n_2, therefore only rough values of n_2 and cos y are needed and K_2h_2 can be determined from the nearest (R_0, T_0) graph to the chosen n_2.

II. *High absorption.*

For high absorption the interference maxima and minima disappear and the influence of interference can be neglected, and we can determine k_2 from the expression

$$k_2 = \frac{\lambda}{4\pi h_2} \ln\left(\frac{1-R_0}{T_0}\right) \qquad \ldots\ldots(15)$$

and n_2 from the expression for the reflection from the front surface of an absorbing medium, viz.

$$R_0 = [(n_2-1)^2 + k_2{}^2]/[(n_2+1)^2 + k_2{}^2]. \qquad \ldots\ldots(16)$$

Again it is convenient to plot R_0 against n_2 for various values of k_2, as in figure 4.

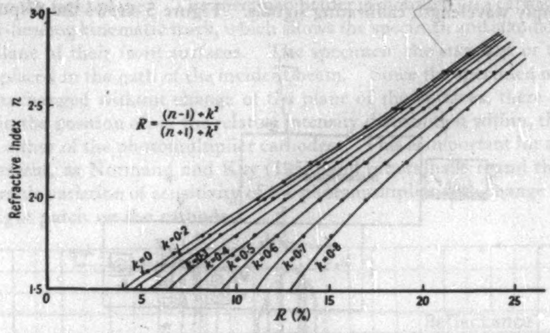


Figure 4. Front surface reflection curves for various values of refractive index and extinction coefficient.

The value of n_2 so obtained may be used to determine a more refined value of k_2 from (Kuwabara 1954)

$$k_2 = \frac{\lambda}{4\pi h_2}\left[\ln\frac{1}{T_0} + \ln\left\{\frac{16 m_3(n_2{}^2 + k_2{}^2)}{[(n_2+1)^2 + k_2{}^2][(n_2+n_3)^2 + k_2{}^2]}\right\}\right]. \qquad \ldots\ldots(17)$$

However, when $T_0 < 5\%$, the original unrefined values are as good as is warranted by the experimental accuracy. Then the procedure in this region is

1. Determine k_2 from equation (15).
2. Determine n_2 from the graphs of figure 4.
3. If necessary, refine k_2 by use of (17).
4. Redetermine n_2 from figure 4.

The region of intermediate absorption in the long wavelength tail and near the absorption edge requires the full treatment; in the absence of extrema, a guess at n_2, based on the results for the high and low absorption regions, is made, and refined as already outlined : then K_2 is obtained.

By these methods, as many points as desired may be determined on curves of n_2 and k_2 against wavelength, all the data being derived from one film. This is important if, for example, the variation of these quantities with film thickness is to be determined. In contrast, the methods used by Hass and Saltzberg and Brattain and Briggs require a range of film thicknesses to obtain n_2 and k_2.

It is to be noted that we have plotted curves of R_0 against T_0 for one value of n_2 only, and error might be expected from this source. However, it provides a good approximation in the visible and near ultra-violet regions, and this is all that is really required, since, as soon as absorption is appreciable, the curves of figure 4 are used in preference to the (R_0, T_0) curves.

§ 3. Experimental

All the measurements were made using a Hilger–Müller quartz double monochromator to which was attached suitable sources and arrangements to measure the reflectivity of the slides of fused silica coated with ZnS. The Hilger–Müller monochromator has been modified to scan a wavelength range automatically and to supply wavelength calibrating signals. Figure 5 shows the disposition of the apparatus. Light from a tungsten or hydrogen lamp L is focused by a spherical concave mirror M_1 on to the entrance slit S_1 of the monochromator. A rigid cast aluminium platform with a machined top surface is bolted to the output side of the monochromator, forming a table on which the optical components are mounted. Light of the selected wavelength emerging from the exit slit S_2 is reflected by the plane mirror M_2 on to the spherical concave mirror M_3, which focuses it at F_T, a few centimetres beyond the plane XX of the specimens Z and V in the specimen holder H. The specimens are orientated so that the monochromatic beam of light is incident at an angle of about 4° or less.

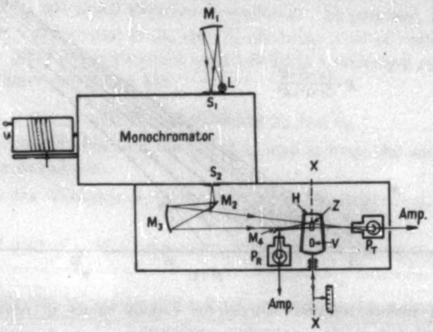


Figure 5. Optical apparatus. L, light source ; M_1, M_3, spherical aluminized mirrors ; M_2, M_4, plane aluminized mirrors ; S_1, S_2, entrance and exit slits of monochromator ; P_R, P_T, reflection and transmission photomultipliers respectively ; H, low temperature cell ; Z, specimen layer evaporated on fused silica ; V, blank fused silica standard ; XX, kinematic slide.

The fraction of the light reflected by the specimen comes to a focus at F_R very close to the edge of the incident cone of light, and is collected by the plane mirror M_4 and reflected on to the photomultiplier P_R. Mirror M_4 is as close as is practicable to the edge of the incident beam, in order to keep the angle of incidence of light on the specimen as small as possible. The specimen, consisting of an evaporated layer of zinc sulphide on a fused silica substrate together with a standard consisting of a similar slip of uncoated fused silica, is mounted in an evacuable specimen holder which can be cooled by liquid air for low temperatures. The specimen and standard are pressed to the ground surface of the edge of a brass frame, with a space between them. This frame is fixed to a metal stem which protrudes through a Dewar flask into a reservoir filled with liquid air when low specimen temperatures are required. Polished fused silica plates are waxed in the walls of the vacuum jacket, through which the light beam can pass. These plates are inclined to the plane of the specimens and to one another, so that stray light reflected from their surfaces does not enter either of the photomultipliers. Suitable shields are placed to trap the strongest such reflected beams. The Dewar, with its metal stem and the specimens on the frame, is fitted to the vacuum cover by the conical ground glass joint, which permits rotation of the specimens to bring them into the right plane. The specimen holder is mounted on a carriage, moving on a ball-bearing kinematic track, which allows the specimen and standard to move in the plane of their front surfaces. The specimen, the standard or the air gap may be placed in the path of the incident beam. Since the specimen or standard can be exchanged without change of the plane of the surfaces, there will be no change in the position of, or the relative intensity distribution within, the patch of light on either of the photomultiplier cathodes. This is important for accuracy of measurement, as Normand and Kay (1952) and others have found that there is considerable variation of sensitivity of the photomultiplier with change in position of the light patch on the cathode.

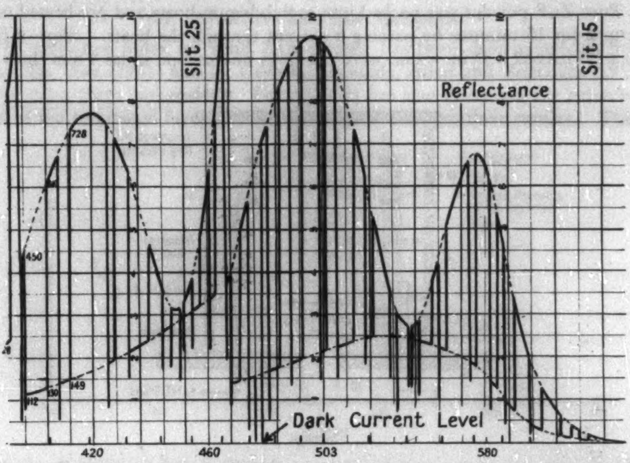


Figure 6. Portion of record of reflection using ' chopping ' technique, showing relative reflections from the specimen and the standard plotted as a function of wavelength.

Both photomultipliers are RCA type IP28 with which measurements can be made down to 2200 mμ. The high-tension supply is obtained from a battery of dry cells, and the output current from the multipliers is amplified in a d.c. amplifier similar to that described by McDonald and Harris (1952). The resulting output is fed into a Leeds and Northrup AZAR Speedomax recorder.

The transmissivity T of the specimen is determined from the ratio of the signal on P_T obtained with the specimen in the beam (I_T) to that with the air gap in the beam (I_0). The reflectivity is determined using the reflectivity of the fused silica standard as a comparison. The reflectivity of the front or back surfaces of the fused silica is given by the Fresnel formula

$$R_0 = (n_3 - 1)^2/(n_3 + 1)^2 \qquad \ldots\ldots(18)$$

and is, of course, a function of wavelength owing to the dispersion of the silica. When allowance is made for the reflections from the front and back surfaces as given above, and the multiple reflections between them, the total reflectivity of the silica slip is given by

$$R = 2R_0/(1 + R_0). \qquad \ldots\ldots(19)$$

The two signals I_R and I_S, obtained on photomultiplier P_R from the specimen and the standard respectively, are both recorded on the one chart by alternatively placing specimen and standard in the beam. Preset stops enable this to be done quickly. Figure 6 shows portion of a typical record derived from a zinc sulphide specimen. The complete curves for I_R and I_S are obtained by interpolation as shown by the dotted lines. The same technique is used to record the ratio between the transmitted and incident radiation. It is intended that the apparatus will be further modified to make the process fully automatic.

§ 4. SPECIMEN PREPARATION

Pure ZnS powder was packed into molybdenum boats and pre-heated under vacuum for 15 minutes by passing a current through the boats. Slips of fused silica ('Vitreosil') 1½ in. × ½ in. were cut from microscope slides 1 mm thick.

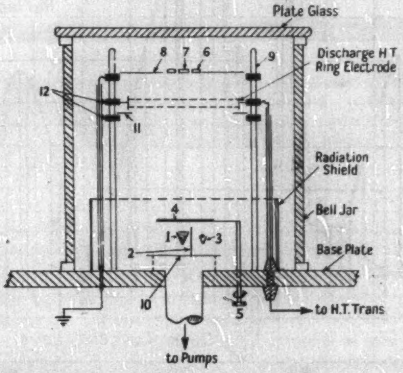


Figure 7. Vacuum evaporation apparatus. 1. Molybdenum boat with ZnS. 2. Shield, 3. Tungsten basket with silver. 4. Evaporation shield. 5. Rotatable joint. 6. Plate glass. 7. Specimen blank. 8. Metal plate. 9. Glass columns. 10. Pump baffle. 11. Discharge electrode shield. 12. High tension ring electrode.

The slips were thoroughly cleaned, placed 25 cm away from the evaporation boat, and vertically above it, as shown in figure 7. A thin polished glass slip was placed immediately beside the specimen slip, for subsequent thickness measurements. A baffle shielded the blanks from the initial vapour stream but was removed when the boat assumed full working temperature. Evaporation rates of the order 100–500 Å per minute were used, in order to avoid films absorbing in the visible region (Rood 1951, Polster 1952). The pressure during evaporation, as measured on ion and Philips gauges was about 5×10^{-5} mm Hg.

Measurements were made on some slips a few hours after evaporation and others were placed in a desiccator for several days before their characteristics were measured.

§ 5. MEASUREMENT OF FILM THICKNESS

The thickness of the ZnS layers was measured by a multiple-beam interferometry technique. A small slab of plate glass, $1\frac{1}{2}$ in. × $\frac{1}{2}$ in. × $\frac{1}{8}$ in., was placed beside the fused silica slip on which ZnS was to be deposited. A portion of this plate glass slip was shielded from the ZnS stream by a brass foil paddle which pressed closely to the glass surface, but which could be removed under vacuum, and the step silvered. The displacement of multiple-beam fringes at the step was used to determine the step height, and hence the thickness of the ZnS.

The consistency of results suggests that the error in measurement of step height, with contributions from a number of sources, was about 20 Å, i.e. 2% of a layer 1000 Å thick. From the geometry of the evaporator, the difference in thickness of the layers on the glass and silica slides is less than 0·1%.

§ 6. RESULTS

6.1. *Refractive Index and Absorption*

The refractive index for solid ZnS crystals has been determined by Brun (1930), De Vore (1951) and by Czyzak, Reynolds, Allen and Reynolds (1954), for the cubic form (sphalerite). Piper (1953) has also discussed the refractive index of the hexagonal form (wurtzite). Figure 8 shows their results. There should be little difference between the refractive indices of the two phases, as they have the same ionic components and almost the same configuration. It is seen that Piper's figures for the hexagonal form agree closely with those for the cubic form in the region where they overlap. No account has been taken of the double refraction in wurtzite, as the difference between the two refractive indices is small.

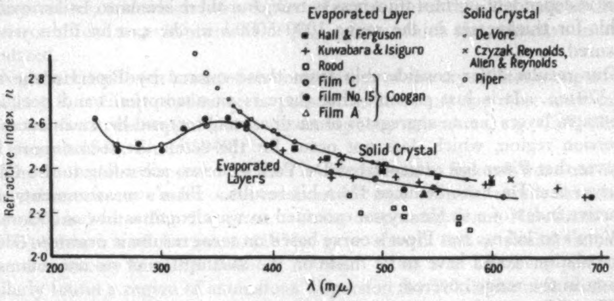


Figure 8. Refractive index of ZnS for crystal and evaporated specimens at room temperature.

It is not clear from Piper's paper whether the dispersion curve he adduces is a result of extrapolating de Vore's results, which do not extend beyond 365 mμ, further into the ultra-violet, or whether it is based on his own experimental measurements of transmission. If the latter is the case, it would appear that no account has been taken of absorption in the crystal, as the expressions derived by Barnes and Czerny (1931) for a thin lamina of non-absorbing material were used to determine the refractive index.

Our results for the refractive index of evaporated ZnS films at room temperature are also plotted in figure 8. The curve is drawn through the average of the results and the points for three films of thickness 4740, 2130 and 1570 Å are plotted on it. In all, results were obtained from more than 10 films. It is seen that in the visible region the refractive index for the films is nearly equal to that for the solid, being 97·5% of the value for the solid at 560 mμ. Also plotted on the same graph are the results for the refractive index of thin evaporated films of ZnS obtained by Kuwabara and Isiguro (1952), by Rood (1951), and by Hall and Ferguson (1955 a, b), in the visible region of the spectrum.

There is little difference between Kuwabara and Isiguro's results and ours, except that their average curve is slightly higher. The results of Hall and Ferguson are slightly higher again, but still lower than the curve for the solid. The results obtained by Rood are considerably lower than the others. Possible reasons for the discrepancy between Rood's results and the other two are (i) that the pressure at which his films were deposited was higher than in the other cases (he mentions pressures greater than 10^{-4} mm), (ii) that the method of determining the refractive index from the value of $R_{\max}$ leads to error, due to the effect of a thin surface layer of lower refractive index overlaying the main film, as is discussed in more detail elsewhere. It is probable that there are contributions from both sources. It is possible that the suggestion of Kuwabara and Isiguro that the refractive index is dependent on film thickness is true, but there seems to be no evidence of this for thicknesses in the range 1000–5000 Å in the case of films we have measured.

Our results differ considerably from those quoted by Piper in the range 330–370 mμ. It is just possible that there is an absorption band peculiar to evaporated layers (i.e. to aggregates of small crystallites) and hence an anomalous dispersion region, which does not occur for the solid. It seems more likely however that Piper has extrapolated De Vore's curves, according to the Cauchy equation that De Vore deduced from his results. Brun's measurements of the refractive index of a single crystal extended in wavelength as low as 393 mμ and De Vore's to 365 mμ but Piper's curve based on these results is drawn to 330 mμ. Extrapolation would have to be made on the assumption of no anomalous dispersion in the range covered.

The difference in the absolute value of the refractive index for solid and evaporated films distant from the absorption bands seems most likely to be due to the state of aggregation of the evaporated films. There is strong evidence that films of evaporated ionic solids are formed of small crystallites of roughly 100–200 Å edge, and that there are small voids between them. With the exception of $CaSi0_3$ (Rood 1949), the apparent refractive index of evaporated films obtained by a number of workers is lower than that for the bulk material. A selection of results for evaporated ionic solids is shown in the table.

It is possible that the higher refractive index for the $CaSi0_3$ film as obtained

Material	**Investigators**	**Refractive Index**		
		Film	**Bulk**	**λ** (mμ)
ZnS	Hammer (1943)	~2·34	2·42	500
	Hermanson (1951)	2·26	2·42	visible
	Rood (1951)	2·22[a]		510
	Kuwabara and Isiguro (1952)	2·37	2·42	500
	Hall and Ferguson (1955 a)	2·39	2·42	500
	Coogan, present paper	2·34	2·42	500
CdS	Gottesman and Ferguson (1954)	2·28[b] 2·21[c] 2·13[d]	O 2·38 E 2·41	750
	Kuwabara (1954)	2·38	O 2·38 E 2·41	750
	Hall and Ferguson (1955 b)	2·42		750
CaF_2	Schulz (1949)	1·29	1·43	
	Bannon (1945)	1·26[e]–1·40[f]	1·43	590
	Abeles (1950)	1·258–1·278[g]	1·43	590
TiO_2	Hass (1955)		O 2·61 E 2·90	590
	Hass and Bradford (1954)	2·33	O 2·61 E 2·90	590
	Banning (1947)	2·2[h]–2·9[i]		visible
$CaSiO_3$	Rood (1949)	1·68	1·61–1·65	590

Remarks: [a] evap. at 100Å/min. Rose after 2 weeks to 2·37. [b,c,d] evap. at 900, 150 and 60 Å/min resp.

[e,f] substrate temp. 20°c, 250°c resp. [g] after aging 2 days. [h,i] deposited at 20°c and 200°c resp.

by Rood may be due to a thin surface layer, in which water is adsorbed between the crystallites, and which has a *higher* refractive index than the bulk of the film. This would lead to enhancement of R_{max}, from which the refractive index was calculated.

No check has been made of the orientation of the crystallites composing the film, but it is likely that some is present as it occurs for most ionic crystals deposited from the vapour phase on amorphous substrates, e.g. CaF_2 (Bannon and Coogan 1949, Shulz 1949).

The anomalous dispersion of ZnS is clearly seen in figure 8 from our results. The peak in refractive index occurs at about 350 mμ. Similar results have recently been obtained for other thin films. Using evaporated films, Kuwabara (1954), and Gottesman and Ferguson (1954), gave shown that a similar anomalous dispersion occurs for CdS at about 500 mμ. Hass and Saltzberg (1954) have similarly found a region of anomalous dispersion for Si0 in the ultra-violet.

Figure 9 shows the refractive index and the absorption curve, as far as it could be determined on this specimen, for a layer 4740 Å thick. The effect of decreasing the temperature to that of liquid air is to 'sharpen' the anomalous dispersion curve considerably though there is very little effect in the visible region. The curves of *R* and *T* plotted against wavelength for this film at room temperature and at the temperature of liquid air are shown in figure 10.

The extinction coefficient k_2 as obtained from a number of films at 20°c and −186°c is shown in figure 11. One can see a suggestion of an unresolved absorption peak at about 330 mμ in the 20°c curve from the kink in the curve

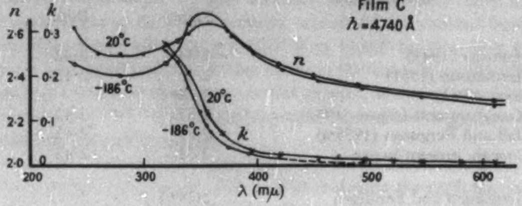


Figure 9. Refractive index and extinction coefficient for ZnS film 4740 Å thick at 20°c and −186°c.

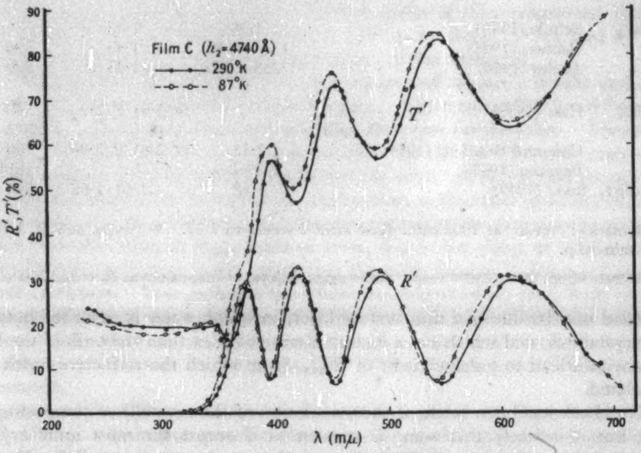


Figure 10. Reflection and transmission curves for ZnS film 4740 Å thick at room temperature and liquid air temperature.

there, which is sharpened at −186°c. This coincides with the centre of the anomalous dispersion region. The absorption at 330 mμ is of the order 10^5 cm^{-1}, showing that the process is one concerned with the crystal matrix and not with a defect absorption process.

The long wavelength tail extends across the visible region in contrast with the much shorter tail shown by the solid. The value of k_2 for $\lambda > 400$ mμ is plotted on an increased scale, for a number of films, and the results of Kuwabara and Isiguro are included for comparison. The shape of the long wavelength tail varies from film to film, as does its absolute magnitude. Gottesman and Ferguson (1954) obtain similar values for k_2 in the long wavelength tail of evaporated CdS, and here, too, the absolute magnitude is very high.

It is almost certain that the crystallites grow by the spiral dislocation mechanism (Burton, Cabrera and Frank 1951). There would thus be a large proportion of each crystallite in which the strain associated with the dislocation

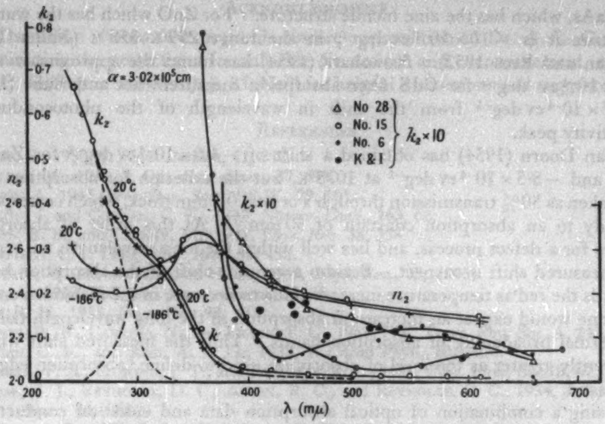


Figure 11. Extinction coefficient for evaporated ZnS layers.

is significant in relaxing the crystal absorption selection rules (Seitz 1951). The absorption introduced thereby would be highly dependent on the average crystal size, and rate of growth, thickness, etc.

It is possible, too, that some of this 'absorption' may be due to scattering of light by the crystallites, which would have the same effect as absorption in the form of measurement employed.

It is probable that the scarcely resolved absorption peak at about 330 mμ in our results (figure 11), which appears to be associated with the anomalous dispersion, is due to an exciton absorption band. This will be discussed further elsewhere.

Our results agree well with those obtained by Mollwo (1944) for absorption in ZnS at room temperature, except in the long wavelength tail. They might well be expected to differ here for several reasons, viz. (i) Mollwo has not taken the effect of interference into account, and (ii) the absorption here is very sensitive to the conditions of formation of the ZnS layer. Shalimova (1951) obtained peaks at 334 mμ for a number of activated ZnS films, but not for pure ZnS, where peaks were resolved at 284 mμ and 313 mμ. He attributes the peak at 334 mμ to excess zinc atoms, but it is clear from our measurements that the absolute value of the extinction coefficient at 334 mμ is far too high to be accounted for by defect absorption processes.

6.2. *Absorption Edge Shift*

From figure 9 it can be seen that the shift in the absorption edge of ZnS, determined from the wavelength for which $k_2 = 0{\cdot}1$, is about 8 mμ on changing the temperature from 97°K to 293°K, or about $-3{\cdot}8\times10^{-4}$ ev deg^{-1}. This shift in the absorption edge towards the red for increasing temperature seems to be common to most ionic solids. For example Cunnell, Edmond and Richards (1954) have recorded a variation of -5×10^{-4} ev deg^{-1} in the range 97°K–293°K for CaAs, which has the zinc blende structure. For ZnO which has the wurtzite structure it is -10×10^{-4} ev deg^{-1}, in the range 299°K–358°K (Miller 1951, Coogan and Rees 1952). Kuwabara (1954) has found the approximate value -5×10^{-4} ev deg^{-1} for CdS from absorption measurements and Bube (1955) $-5{\cdot}2 \times 10^{-4}$ ev deg^{-1} from the shift in wavelength of the photoconductive sensitivity peak.

Van Doorn (1954) has obtained a shift of $-4{\cdot}6 \times 10^{-4}$ ev deg^{-1} for ZnS at 77°K and $-8{\cdot}5 \times 10^{-4}$ ev deg^{-1} at 1073°K, but the criterion for absorption edge was taken as 50% transmission through a crystal 0·2 mm thick, which corresponds roughly to an absorption constant of 50 mm^{-1}. As this order of absorption occurs for a defect process, and lies well within the long wavelength tail region, the measured shift is suspect. Besides a genuine shift of the absorption bands towards the red as temperature increases, due to decrease in the forbidden energy gap, one would expect an increase in absorption in the long wavelength tail due to thermal broadening of absorption bands. Thus the measured shift will be apparently greater as the level of absorption used to define 'absorption edge' is decreased.

Using a combination of optical absorption data and electrical conductivity data, Piper (1953) has obtained the value -3×10^{-4} ev deg^{-1}.

The spread in the values cited above for ZnS band gap shift is probably mainly due to differences in definition of the band gap as measured experimentally. We shall discuss the theoretical aspects of thermal band gap shift elsewhere.

6.3. *Exposure Effects*

The low-temperature cell in which the specimens were normally placed could be evacuated. It was found that when a specimen was placed in the cell *in vacuo*, and the reflectivity measured, the maxima in the reflectivity curve were greater than the values obtained in air. This effect was reversible. Thus an increase of reflectivity from 34·6% to 35·8% was found in the case of a typical specimen when the cell was evacuated. The effect was not instantaneous, but required about 12 hours to reach equilibrium in the new environment.

§ 7. CONCLUSION

The absorption spectrum of ZnS as obtained contains a band with its peak at about 330 mμ which is more apparent at liquid air temperatures. Coinciding with this peak there is a region of anomalous dispersion which can be attributed to it. This band we believe to be an exciton band.

The method of abstracting the optical constants of thin films, outlined above, is applicable to most solids, and in particular to those of high refractive index. There has been some speculation and discussion lately regarding the nature and position of exciton bands (Seitz 1954). It seems that a valuable contribution could be made to solid state physics if a survey were made of the optical properties of some of the prominent ionic solids, in particular measuring the ***absolute*** value of the absorption constant and refractive index versus wavelength. There would then be a firmer basis for interpretation, as the early results are qualitative, and do not give the absolute magnitudes of the absorption constant (e.g. Fesefeld 1930), nor have the dispersion curves been obtained previously.

ACKNOWLEDGMENT

The writer wishes to express his sincere appreciation to Dr. A. L. G. Rees for initiating the project and for valuable discussions and encouragement, and to Mr. C. Billington and Mr. J. J. McNeill for helpful discussions and suggestions.

REFERENCES

ABELES, F., 1950, *J. Phys. Radium*, **11**, 310.
BANNING, M., 1947, *J. Opt. Soc. Amer.*, **37**, 688.
BANNON, J., 1945, *Proc. Roy. Soc. N.S.W.*, **79**, 101.
BANNON, J. and COOGAN, C. K., 1949, *Nature, Lond.*, **163**, 62.
BARNES, R. B., and CZERNY, M., 1931, *Phys. Rev.* **38**, 338.
BRASEFIELD, C. J. 1940, *Phys. Rev.*, **58**, 436.
BRATTAIN, W. H., and BRIGGS, H. B., 1949, *Phys. Rev.*, **75**, 1705.
BRUN, A., 1930, *Bull. Soc. franc. mineral*, **53**, 35.
BUBE, R. H., 1955, *Phys. Rev.* **98**, 431.
BURTON, W. R., CABRERA, N., and FRANK, F. C., 1951, *Phil. Trans. Roy. Soc.* A, **243**, 299.
COOGAN, C. K., and REES, A. L. G., 1952, *J. Chem. Phys.*, **20**, 1650.
CUNNELL, F. A., EDMOND, J. T., and RICHARDS, J. L., 1954, *Proc. Phys. Soc.* B, **67**, 848.
CZYZAK, S. J., REYNOLDS, D. C., ALLEN, R. C., and REYNOLDS, C. C., 1954, *J. Opt. Soc. Amer.*, **44**, 864.
DE VORE, J. R., 1951, *J. Opt. Soc. Amer.*, **41**, 416.
FESEFELD, H., 1930, *Z. Phys.*, **64**, 741.
GISOLF, J. H., 1939, *Physica*, **6**, 84.
GOTTESMAN, J., and FERGUSON, W. F. C., 1954, *J. Opt. Soc. Amer.*, **44**, 368.
HALL, J. F., and FERGUSON, W. F. C., 1955 a, *J. Opt. Soc. Amer.*, **45**, 74 ; 1955 b, *Ibid.*, **45**, 714.
HAMMER, K., 1943, *Z. tech. Phys.*, **24**, 169.
HASS, G., 1955, *J. Opt. Soc. Amer.*, **45**, 945.
HASS, G., and BRADFORD, A. B., 1954, *J. Opt. Soc. Amer.*, **44**, 810.
HASS, G., and SALTZBERG, C. O., 1954, *J. Opt. Soc. Amer.*, **44**, 181.
HEAVENS, O. S., 1951, *Proc. Phys. Soc.* B, **64**, 419.
HERMANSON, A., 1951, *Nature, Lond.*, **167**, 104.
KUWABARA, G., 1954, *J. Phys. Soc., Japan*, **9**, 97.
KUWABARA, G., and ISIGURO, K., 1952, *J. Phys. Soc., Japan*, **7**, 72.
McDONALD, K. L., and HARRIS, F. S., 1952, *J. Opt. Soc. Amer.*, **42**, 321.
MILLER, P. H., 1951, *Semi-Conducting Materials*, Ed. H. K. Henisch (London : Butterworths Scientific Publications), p. 172.
MOLLWO, E., 1944, *Reichsber. Phys.*, **1**, 1.
MOONEY, R. L., 1945, *J. Opt. Soc. Amer.*, **35**, 574.
NORMAND, C. W. B., and KAY, R. H., 1952, *J. Sci. Instrum.*, **29**, 33.
PIPER, W. W., 1953, *Phys. Rev.*, **92**, 23.
POLSTER, H. D., 1952, *J. Opt. Soc. Amer.*, **42**, 21.
ROOD, J. L., 1949, *J. Opt. Soc. Amer.*, **39**, 854 ; 1951, *Ibid.*, **41**, 201.
SCHULZ, L. G., 1949, *Phys. Rev.*, **75**, 345 (A).
SEITZ, F., 1940, *Modern Theory of Solids* (New York : McGraw-Hill), p. 449 ; 1951, *Rev. Mod. Phys.*, **23**, 328 ; 1954, *Ibid.*, **26**, 7.
SHALIMOVA, K. V., 1951, *Dokl. Akad. Nauk. S.S.S.R.*, **80**, 587.
VAN DOORN, C. Z., 1954, *Physica*, **20**, 1155.
WRIGHT, D. A., 1948, *Proc. Phys. Soc.*, **60**, 22.

The Temperature Coefficient of Reflectivity of Nickel

By L. WARD[†]
University College of Ghana, Accra

MS. received 19*th March* 1957

Abstract. The variation with temperature of the reflecting power of nickel has been measured by a null method at five wavelengths in the near infra-red region of the spectrum. The results are in satisfactory agreement with those calculated from previous work on emissivity. There was no conclusive evidence for a significant change in reflectivity at the Curie point.

§ 1. Introduction

This work was suggested by the results of an investigation into the emissivities of iron, nickel and cobalt when the temperature coefficients of emissivity were measured above 700°c at wavelengths between 1 μ and 3 μ (Ward 1956). At the same time, in the cases of iron and cobalt, it was observed that there were relatively sharp discontinuities in the temperature coefficients in the regions of the respective Curie points. A study of the reflecting power of nickel afforded an opportunity both of measuring the temperature coefficients at low temperatures and of observing possible changes at the Curie point (370°c) which was inaccessible in the previous work.

In the case of thick metals, the interconversion of emissivity E and reflectivity R is easily effected by the relation $R=1-E$. The measurement of reflectivity offers two considerable advantages over that of emissivity. Firstly, there is no necessity to produce standard black body conditions, and secondly, the overall response of the apparatus does not alter greatly with the specimen temperature, as the temperature variation of reflectivity is a second order effect. The intensity of the measured radiation is a function of the source temperature, variations in which can be eliminated by using a double beam null method. Against these may be set two disadvantages : (*a*) as the reflectivity in the near infra-red is between 2 and 4 times greater than the emissivity, the temperature coefficient of reflectivity $(1/R)(dR/dT)$, is only between a half and a quarter that of emissivity $(1/E)(dE/dT)$ and (*b*) the reflecting surfaces must be accurately plane and highly polished to produce sharp images. A long and tedious process is involved in (*b*) although the effect of (*a*) can be minimized by a multiple reflections method.

§ 2. Apparatus

As the main emphasis was on studying the variation in reflectivity with temperature, a comparative method of making measurements was employed. A double beam arrangement of mirrors was used in conjunction with an infra-red spectrometer and a lead sulphide cell. A pair of reflecting nickel surfaces was introduced into each beam, one pair being heated and the other kept cool.

[†] Now at College of Technology, Birmingham.

The complete layout of the apparatus is shown in figure 1. The source of radiation was a 12 v 36 w car headlamp bulb placed at the foci of the two collimating mirrors. The four nickel mirrors were mounted inside a vacuum-tight chamber, the beams entering and leaving through quartz windows. The two beams were focused by the telescope mirrors and, by means of a system of plane mirrors, entered the spectrometer along the same path. One mirror, having a reflectivity of about one third at 1 μ, reflected the first beam directly into the spectrometer. The second beam passed through this mirror but was reflected back into the spectrometer by a fully-silvered mirror directly behind. After the double transmission through the partially silvered mirror the second beam was of roughly the same intensity as the first.

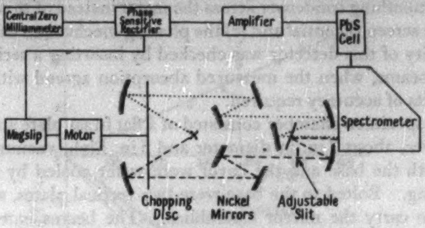

Central Zero Milliammeter
Phase Sensitive Rectifier
Amplifier
PbS Cell
Spectrometer
Magslip
Motor
Chopping Disc
Nickel Mirrors
Adjustable Slit

Figure 1. General layout of the apparatus.

The spectral radiation fell on a lead sulphide cell coupled to a phase sensitive detector. The beams were interrupted, at a frequency of 22 c/s, by a rotating disc having a single aperture 1 in. square, so that alternate beams reached the cell. The motor driving the disc was also directly coupled to a magslip which generated a sinusoidal voltage with frequency and phase locked to that of the disc. The relative phase of this signal and that from the PbS cell could be adjusted by rotating the stator of the magslip.

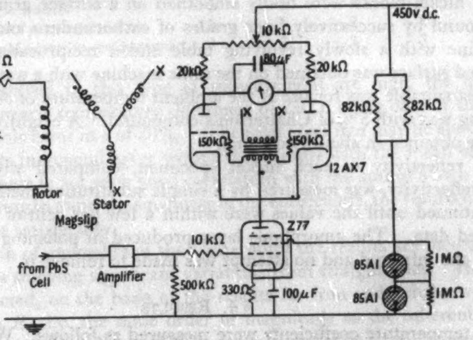

450v d.c.
10 kΩ
20kΩ
80μF
20 kΩ
25 Ω
X
12v d.c.
82kΩ
82kΩ
X
X
150kΩ
150kΩ
12 AX7
Rotor
X
Stator
Magslip
Z77
10 kΩ
from PbS Cell
Amplifier
500 kΩ
330Ω
100μF
85A1
1MΩ
85A1
1MΩ

Figure 2. Circuit of phase sensitive detector.

The operation of the phase sensitive detector may be seen from the circuit diagram shown in figure 2. The signals representing the intensities of the two beams were amplified and passed to the cathode of a double triode. The magslip voltage was used to switch them in turn to each triode, a null reading on the centre-zero milliammeter being obtained by adjusting a calibrated slit placed in one of the beams. The circuit is similar to one used by Schuster (1951), the mean differences being (*a*) the smoothing condenser across the meter instead of to earth. (*b*) neons to stabilize the screen potential and (*c*) the phasing mechanism described above.

The linearity of this detector was checked by inserting a series of glass plates in one of the beams, when the measured absorption agreed with that calculated within the limits of accuracy required.

The actual specimen chamber consisted of a flat brass plate and a demountable cylindrical cover, about 7 in. in diameter and 5 in. high, vacuum sealed with an 'O' ring. Both the base and the cover were water cooled by means of spirals of copper tubing. Bolted to the base were two vertical plates, aligned along the optical path, to carry the mirror assemblies. The beams entered and left the chamber by four ports in the cover each closed by a quartz window and sealed with an 'O' ring.

Each mirror, 1 in. square by ¼ in. thick, was heated by a nichrome element sandwiched in mica and clamped by four bolts between the mirror and a backing plate of ⅛ in. steel. A hole was drilled in the nickel block parallel to the reflecting face, and a single bolt through this was used to attach the mirror to the vertical plate. In addition to providing a convenient means of adjustment, this method of clamping minimized the displacement of the image due to distortion on heating.

The temperature of the mirror face was measured with a platinum–platinum-rhodium thermocouple fastened beneath one of the clamping bolts. Heating was carried out in a vacuum of 0·03 mm Hg and temperatures up to 600°c could be obtained by passing 5 A from a Variac through the heating element. There was no evidence of oxidation taking place under these conditions.

§ 3. POLISHING THE SPECIMENS

Polishing the specimens presented one of the major problems of the investigation, and the following technique was found to be the most satisfactory.

The nickel blocks were firstly smoothed on a surface grinder. They were then ground by successively finer grades of carborundum and Sira powder on a machine with a slowly revolving table and a reciprocating arm. Finally, a polished surface was obtained on the same machine with a wax block and rouge. The most suitable wax for use at the ambient temperature of 30°c was a mixture of sealing wax with 5% of Chatterton's compound. A brightly polished surface could be obtained in about two hours.

The reflectivity of each nickel specimen, compared with aluminium of known reflectivity, was measured by a simple substitution method and polishing was continued until the values were within a few per cent of the most reliable published data. The amorphous layer produced in polishing has only a small effect in the infra-red and no attempt was made to remove it.

§ 4. RESULTS

The temperature coefficients were measured as follows. With both pairs of mirrors at room temperature the slit width w was adjusted to give null output. After raising the temperature of one pair of mirrors by T°c the change in slit width d to restore balance was determined. Then $d/w = 2\Delta R/R_0 = 2\alpha T$, R_0 being the room temperature reflectivity and ΔR the change over T°c. Figure 3 shows $\Delta R/R_0$ plotted against T for wavelengths of 1·0, 1·4, 1·8, 2·0 and 2·4 μ, and the table gives the values of temperature coefficients determined from these graphs.

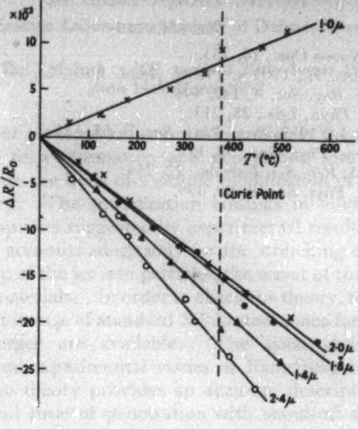


Figure 3. Graph of $\Delta R/R_0$ plotted against T for nickel.

Also shown in this table are the calculated values from (*a*) the author's previous emissivity results and (*b*) the mean results of seven previous investigations (Hagen and Rubens 1903 a, b, Hurst 1933, Cennamo 1939, Price 1947, Lund and Ward 1952, Ward 1956).

	λ (μ)	1·0	1·5	1·8	2·0	2·4
α . 10^5deg.$^{-1}$	Present results	2	−5	−4	−4	−6
	Author's previous results	5	−1	−2	−2	−3
	Mean of seven investigations	3	−3	−3	−3	−4

The errors involved in this type of measurement are high, the quantity $2\Delta R/R_0$ (i.e. d/w) over a temperature difference of 250°c being only of the order of 2% with a possible error in d of 20%. The total error in α may be about 25%. The present results in general agree within these limits with the mean of earlier values, although there is not such good correlation with the author's own results. On the whole, bearing in mind the smallness of the coefficients, the results obtained would appear to be satisfactory.

On the question of anomalies at the Curie point the results are inconclusive, none of the graphs showing significant variations from straight lines. However, the anomaly expected, on the basis of the results for iron and cobalt, would be only about ½% in R_0, i.e. the same order of magnitude as the inherent errors. It is interesting to note that Lowe (1936), measuring the spectral emissivities of nickel, only observed Curie point anomalies above 4·5 μ.

ACKNOWLEDGMENTS

The author wishes to acknowledge the assistance of Mr. E. E. Barratt and Mr. P. W. J. Black of the University College of Ghana in connection with the design and construction of the vacuum and electronic aspects of the apparatus.

REFERENCES

CENNAMO, F., 1939, *Nuovo Cim.*, **16**, 253.
HAGEN, E., and RUBENS, H., 1903 a, *Preuss. Akad. Wiss.*, **13**, 269; 1903 b, *Ibid.*, **19**, 410.
HURST, C., 1933, *Proc. Roy. Soc.* A, **142**, 446.
LOWE, E., 1936, *Ann. Phys.*, *Lpz.*, **25**, 213.
LUND, H., and WARD, L., 1952, *Proc. Phys. Soc.* B, **65**, 535.
PRICE, D. J., 1947, *Proc. Phys. Soc.*, **59**, 131.
SCHUSTER, N. A., 1951, *Rev. Sci. Instrum.*, **22**, 254.
WARD, L., 1956, *Proc. Phys. Soc.* B, **69**, 339.

Penetration by High-Speed Metallic Jets

By SAMPOORAN SINGH

Defence Science Laboratory, Ministry of Defence, New Delhi, India

Communicated by F. C. Auluck ; MS. received 20th August 1956, *and in revised form* 14*th March* 1957

Abstract. This paper presents an attempt to correlate theoretically the various phenomena relating to jet formation and penetration into metallic targets by lined cavity charges on the basis of a simple modification of the non-steady state hydrodynamic theory. The modification consists in adding to the theory a couple of new assumptions suggested by experimental results and it appears that this modified theory accounts adequately for the stretching of the jet as it travels in space, the break-up of the jet into particles, the waver of the jet and the strength of the jet and target materials. In order to check the theory, numerical evaluations have been carried out in case of standard M9Al steel cones for which experimental results of Eichelberger are available. The good quantitative agreement between the published experimental values of Eichelberger and the theoretical results show that the theory provides an accurate description of the variation of depth, velocity and time of penetration with stand-off and nature of target material.

§ 1. INTRODUCTION

THE basic theory of penetration by high-speed metallic jets was developed independently during World War II by Pugh (Birkhoff *et al.* 1948) and by Pack and Evans (1951). The interesting conclusion is that the depth of penetration into a massive target depends only on the length and density of the jet and the density of the target, but is independent of the jet velocity. This result is expressed by a familiar equation:

$$P = L(\lambda\rho_j/\rho_t)^{1/2} \qquad \ldots\ldots(1)$$

where P is the depth of penetration, L the length (assumed constant) of the jet, ρ_j the density of the jet material, ρ_t the density of the target material, and λ a constant which has the value 1 for a continuous jet and 2 for a dispersed particle jet. In a continuous jet the metal occupies the whole volume whereas in a particle jet the metal occupies only a part of the volume. The continuous jet is capable of supporting internal pressure whereas the particle jet is not. Pack and Evans also took the strength of the target into consideration and suggested that

$$P = L(\lambda\rho_j/\rho_t)^{1/2}(1 - k\sigma) \qquad \ldots\ldots(2)$$

where σ is proportional to the initial yield strength of the target material and k is an empirical factor depending on the properties of the jet and the density of the target. The velocity gradient (Birkhoff *et al.* 1948, Eichelberger and Pugh 1952) in the jet causes any element of the jet to elongate as it travels in space and this means, in effect, that the values of L and $\lambda\rho_j$ (and probably of k) must change during the process. Pugh also discussed the break-up and waver in the jet in an approximate manner by assigning empirical values to constants depending upon the velocity gradient and rate of radial spreading.

Recently Singh (1953) calculated the penetration by assuming the jet to be divided into small elements of finite length and equation (2) to hold good for each jet element. The depth of penetration ΔP by a given jet element of length ΔL is

$$\Delta P = \Delta L \frac{U}{V-U}(1-k\sigma) = \Delta L \left(\frac{\lambda\rho_j}{\rho_t}\right)^{1/2}(1-k\sigma) \qquad \ldots\ldots(3)$$

where $\lambda\rho_j$ is the effective density (average mass per unit volume) of the jet element, V is the mean velocity of the jet element, U the time rate of change of the depth of the hole and k an empirical factor for each jet element and the given target. Singh also suggested the addition of a term to the basic equation of the theory that would permit strength effects to be taken into account:

$$\lambda\rho_j(V-U)^2 = \rho_t U^2 + 2\sigma. \qquad \ldots\ldots(4)$$

Combining equations (4) and (3), the term k is evaluated

$$k = \frac{1}{\rho_t V^2}\left[\left(\frac{\rho_t}{\lambda\rho_j}\right)^{1/2} + 1\right]^2. \qquad \ldots\ldots(5)$$

The total penetration is the summation of penetrations by all the jet elements. Recently Eichelberger (1956) suggested to substitute the product $\gamma\rho$ for $\lambda\rho_j$ where ρ is the density of the liner material and all corrections for discontinuities within the jet are made by varying γ. This includes any changes in the mode of transfer of momentum, separation of the particles after the jet break up and waver of the jet due to imperfections in charge or liner. Eichelberger also presented experimental measurements of jet velocity, penetration velocity and time of penetration as functions of depth of penetration by firing lined cavity charges into several types of target material.

The object of the present paper is to correlate theoretically the non-steady state hydrodynamic theory of jet formation and target penetration by lined cavity charges; and is basically an extension of the author's previous work (Singh 1953). Some alterations in ideas concerning the mechanism of penetration by the jet are also presented. In order to check the theory numerical evaluations have been carried out in case of standard M9Al steel cones for which experimental results are available (Eichelberger 1955). The calculated depth, velocity and time of penetration in several types of target material are compared with published experimental results (Eichelberger 1956).

§ 2. MODIFICATIONS TO THE THEORY OF PENETRATION

Let β represent the angle between the collapsing surface of the cone and the axis, V_0 the velocity with which the cone element travels towards the axis, V_j the velocity of the jet element formed and x the length measured from the apex along the axis to the plane of the zonal element in the cone. Eichelberger (1955) determined β as a function of x by means of slug mass measurements and calculated V_0 and V_j for the standard M9A1 steel cones in the standard C.I.T. laboratory charge (Eichelberger and Pugh 1952). These collapse parameters (β, V_0 and V_j as functions of x) are taken as the basis of all numerical calculations. Figure 1 shows the geometrical representation of the formation of a jet element and its penetration in a target. Let A, B, . . . , L represent different elements from the apex to the base in the original conical liner. The length of a finite element in

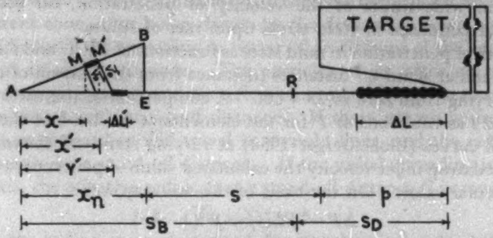


Figure 1. Geometrical representation of the formation of a jet element and its penetration in a target. AMB is the upper half and AE the axis of the original conical liner. $\Delta L'$ represents the length of the jet element that arises from a finite element MM' at the instant t'. $\Delta L'$ stretches as it travels in space ; at R the break-up of the jet element starts, and ΔL represents the length of the jet element when it is about to strike the target. p represents the depth of penetration by the jet elements arising from AM length of the original liner. (The thicknesses of the element MM' and that of jet elements $\Delta L'$ and ΔL are shown in the figure).

the parent cone is fixed by lengths x and x' measured from the apex along the axis to the plane of the given zonal element. For element A, x and x' are 0·52 cm and 0·75 cm respectively. The length of each of the elements B, C, . . . , L is 0·25 cm. Each finite element, when it reaches the axis, gives an element of jet of length $\Delta L'$. This stretches as it travels in space: let ΔL be its length when it is about to impinge on the target. Obviously, the jet elements ΔL_A, ΔL_B, . . . , ΔL_L arise from the elements A, B, . . . , L in the parent cone, respectively. Each element of the jet can also be characterized by a finite velocity of the head V_J and the tail end V_J'. Thus, in principle, each jet element (identified by its velocity) after having travelled a wide variety of distances to the target is studied. We further assume that there is no interaction between adjacent elements of the jet, that is each element of the jet acts independently of the rest.

The expressions for ΔP, k and U are as follows:

$$\Delta P = \Delta L(\gamma\rho/\rho_t)^{1/2}(1-k\sigma), \qquad \ldots\ldots(6)$$

$$k = \frac{1}{\rho_t V^2}\left[\left(\frac{\rho_t}{\gamma\rho}\right)^{1/2}+1\right]^2 \qquad \ldots\ldots(7)$$

and

$$U = V\left[\frac{1}{1+(\rho_t/\gamma\rho)^{1/2}} - \frac{\sigma}{\rho_t V^2}\left(\frac{\rho_t}{\gamma\rho}\right)^{1/2}\right] \qquad \ldots\ldots(8)$$

to the first order in $\sigma/\rho_t V^2$. Let Δt be the time taken for a jet element ΔL to penetrate ΔP; then Δt is given by the expression

$$\Delta t = \Delta P/U. \qquad \ldots\ldots(9)$$

The quantity σ is presumed to be the difference between two quantities σ_t and σ_J which represent the resistance of the target and the jet respectively, to the plastic deformation required by the penetration process (Eichelberger 1956). Since the liner is of mild steel, a value $\sigma_J = 2\times10^9$ dyn cm^{-2} is used throughout for the resistance of the jet to deformation. The factor σ_t is assumed to be some multiple of static yield strength of the target material. When a hole is made dynamically σ_t depends also upon the mean rate of strain. For large strain rates σ_t will not vary between wide limits of the velocity of penetration, on account of the logarithmic dependence of yield stress upon rate of strain.

The depth of penetration in mild steel as functions of V_J, U and t are calculated taking $\sigma = 0$ and at stand-off distances (distance from the bottom of charge to top of target) varying from zero to 25·4 cm. A comparison of the theoretical curves of V_J, U and t as functions of P for the elements at the head of the jet with the experimental curves (Eichelberger 1956) at varying stand-off distances indicates that for a fixed drop in jet velocity the calculated value of penetration is more than the experimental value. On the basis of this we may write

$$\Delta P = \epsilon \Delta L(\gamma\rho/\rho_t)^{1/2}(1 - k\sigma) \qquad \ldots\ldots(10)$$

where ϵ is some empirical factor. ϵ has been found to be 0·5, 0·6, 0·7, 0·8, 0·85, 0·9 and 0·95 for elements ΔL_A to ΔL_G of the jet respectively and unity for elements ΔL_H to ΔL_L. This modification is dictated by the nature of the experimental observations. Probably the need for this empirical factor arises from the fact that (*a*) for the first few elements of the jet the quasi-stationary penetration formula is inadequate, (*b*) the elements at the head of the jet have to start the lateral motion of the target and there are also the surfaces effects, and (*c*) the elements at the head of the jet have a comparatively small amount of material as compared with those at the tail of the jet. An evaluation of $\Delta L'$ and ΔL for the elements at the head of the jet indicates that these remain continuous at very large stand-off distances in spite of the very large strain rates; the actual situation, however, is expected to be considerably complicated.

Each finite element of the jet elongates as it travels in space. There are two cases to consider and both may occur at different stages in the same element of the jet.

Case 1. ($\Delta L \leqslant C\Delta L'$). Each element of the jet is originally a continuous stream. It stretches (like a ductile metal) as it travels in space and its cross-sectional area decreases. As long as an element of the jet is continuous, it is expected to behave as a perfect fluid at the target so that $\gamma = 1$. As the stand-off distance increases each continuous element of the jet has more opportunity to stretch before it reaches the target, and so penetrates more and more efficiently.

Case 2. ($\Delta L > C\Delta L'$). The increasing efficiency of penetration with increasing stand-off distance does not continue indefinitely due to two factors. The first of these is normal break-up, which would be experienced even by a 'perfect' jet. The amount of ductile drawing that each element of the jet will withstand is naturally limited. After it has become elongated to its limits it will break up into discrete particles. We assume that each element of the jet begins to break up into discrete particles when $\Delta L > C\Delta L'$, where C is an empirical factor†, whose value (Eichelberger 1956) is 4 to 5 approximately. Let S_B (figure 1) be the axial distance from the top of the cone to the point up to which the jet element remains continuous (or beyond which the jet element starts breaking into discrete particles). S_B is given by the expression (Singh 1953)

$$S_B \simeq \Delta L'(C-1)\frac{V_J}{V_J - V_J'} + Y' + \Delta L' \qquad \ldots\ldots(11)$$

† The velocity gradient $[=(V_J - V_J')/\Delta L']$ in a jet element which remains continuous can be directly transposed into a strain rate. For various jet elements of length $\Delta L'$, the rate of strain ($\sim 10^5$ sec^{-1} to 2×10^5 sec^{-1}) is approximately the same and this suggests that the term C for most of the jet elements is approximately the same.

where Y' is the axial distance from the top of the cone to the point where an element x' in the parent cone, after being driven by the detonation wave, meets the axis. The second factor is the spreading or waver of an element of the jet and is due to imperfections in charge or cone. The particles of jet, after normal break-up, travel as a diverging cone† and the effective cross-sectional area of an element of the jet increases with stand-off distance. If there are no appreciable forces upon the jet particles they will travel in straight lines, and the radial spreading will vary linearly with stand-off distance. If the radial spreading is symmetrical about the axis, the effective radius at any stand-off distance S is roughly

$$R_j + S_d \tan\phi,$$

where R_j is the radius of an element of the jet when it just starts breaking into particles, S_d the axial distance from the point of normal break-up of an element of the jet to the point of its impingement on the target and ϕ the angle of spread which the particles subtend with the axis. S_d is given by the following expression:

$$\begin{aligned} S_d &= x_n + S + p - S_B \\ &= x_n + S + p - Y' - \Delta L' - \Delta L'(C-1)\left(\frac{V_j}{V_j - V_j'}\right) \quad \ldots\ldots(12) \end{aligned}$$

where x_n is the height of the cone and p the depth of penetration by all elements of the jet which are ahead of the particular element under consideration.

The conservation of mass equation is as follows:

$$\Delta m_j = \pi R_j^2 C \Delta L' \rho = \pi (R_j + S_d \tan\phi)^2 \Delta L \gamma \rho \quad \ldots\ldots(13)$$

where Δm_j is the mass of the jet element. From the above equation the break-up factor γ is given by

$$\gamma = \frac{C\Delta L'}{\Delta L}\left[\frac{R_j}{R_j + S_d \tan\phi}\right]^2 \quad \ldots\ldots(14)$$

where $R_j = (\Delta m_j / \pi C \Delta L' \rho)^{1/2}$‡. Obviously the magnitude of γ depends on the simultaneous effects of break-up and of jet waver and must always be less than unity for a particle jet. It is interesting to point out that this interpretation of γ is in conformity with that suggested by Eichelberger (1956). We further assume that each particle individually penetrates in the manner described by the simple theory for a continuous jet. Substituting the value of γ from equation (14) in equations (10) and (7), we have

$$\Delta P = \epsilon \Delta L \left(\frac{C\Delta L'}{\Delta L}\,\frac{\rho}{\rho_t}\right)^{1/2}\left(\frac{R_j}{R_j + S_d \tan\phi}\right)(1 - k\sigma) \quad \ldots\ldots(15)$$

where

$$k = \frac{1}{\rho_t V^2}\left[\left(\frac{\Delta L}{C\Delta L'}\,\frac{\rho_t}{\rho}\right)^{1/2}\left(\frac{R_j + S_d \tan\phi}{R_j}\right) + 1\right]^2.$$

The terms $\Delta L'$, ΔL, V, R_j and S_d can be evaluated from the collapse parameters

---

† Following Kolsky *et al.* (1949), the travel of jets in air from shaped charges having 45° conical liners of steel and copper, and of 1½ in. to 2½ in. calibre were photographed by opening the shutter of the camera in complete darkness before firing the charges (Singh 1954). The luminescent beam is in the form of several straight tracks confined to a very narrow cone, which is observed better on the original negatives of copper jets. Measurements of the diameter of the luminous beam at two places (at very large stand-off distance) indicate that the angle of spread ϕ, which the particles subtend with the axis, is of the order of a degree only.

‡ The calculated R_j increases from 0·05 cm to 0·11 cm from the head to the tail end of the jet and these values agree very well with that determined from the flash radiographs.

(β, V_0 and V_j as functions of x) of a given cone. The magnitude of ϕ is of the order of a degree only, but its precise value for a given jet element is not known.

If the radial spreading is symmetrical about the axis, one would expect ϕ to be the same at all stand-off distances. The calculated depths of penetration by equation (15) taking ϕ to be the same at all stand-off distances (e.g. $\sim 1°$) do not agree with the experimental ones. This perhaps suggests that the radial spreading may be somewhat non-symmetrical and may be different for different jet elements. However, due to the difficult nature of the theoretical treatment on these lines, we have instead assumed the radial spreading to be symmetrical about the axis and ϕ to vary with stand-off distance, but to be the same for different jet elements at a given stand-off distance. The angle of spread ϕ has been assigned values such that the depths of penetration at different stand-off distances S calculated by equation (15) agree with the experimental data. These are as follows:

S (cm)	0·00	2·54	5·08	7·62	10·16	12·70	15·24	17·78	20·32	22·86	25·40
ϕ (deg)	0·00	0·02	0·06	0·10	0·14	0·18	0·21	0·23	0·25	0·27	0·29

Justification of the above view is to be found from the fact that the calculated depth of penetration, velocity of penetration and time of penetration at several stand-off distances and in several types of target material (see § 3) are in close agreement with the published experimental results (Eichelberger 1956).

§ 3. DISCUSSION OF RESULTS

3.1. *Effect of Stand-off Distance*

Eichelberger (1956) claimed that for mild steel targets, the factor $\sigma(=\sigma_t-\sigma_j)$ is equal to zero. In preliminary analysis, the depth of penetration was calculated taking $\sigma=0$ and $C=4{\cdot}5$. However we found poor agreement between the theoretical and experimental curves (P as function of V_j, U and t) if we put $\sigma=0$; this suggests that even a mild steel target offers resistance far greater than its static yield strength. A reasonably good fit is obtained by taking σ to be 12×10^9 dyn cm^{-2} i.e. σ_t about seven times the static yield strength and σ_j approximately equal to the static yield strength of mild steel. The value of σ is in agreement with that suggested by Evans and Pack (1951) and experimentally determined by Bishop, Hill and Mott (1945) at low rates of strain.

The calculated depths of penetration in mild steel targets (taking $C=4{\cdot}5$, $\sigma=12\times10^9$ dyn cm^{-2}) at varying stand-off distances are shown in figure 2. The

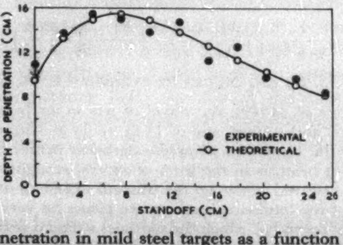


Figure 2. Depth of penetration in mild steel targets as a function of stand-off distance by jets squirted from standard M9A1 steel cones in the standard C.I.T. laboratory charge. The experimental curve is also shown.

experimental average depths of penetration in mild steel at varying stand-off distances are also shown in the same figure. The depths of penetration by different elements of the jet at varying stand-off distances are shown in figure 3.

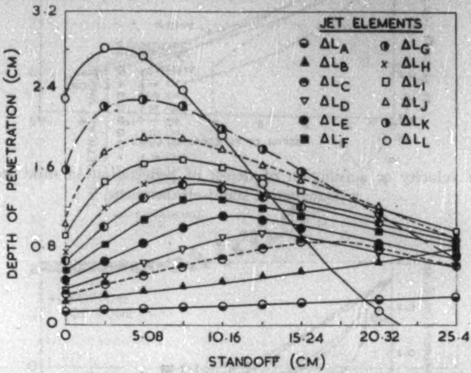


Figure 3. Depth of penetration into mild steel by different elements of the jet for several representative stand-off distances.

This, in principle, gives the contribution to penetration by each jet element (identified by its velocity) after it has travelled a wide variety of distances to the target. The penetrations by elements ΔL_{A} and ΔL_{B} at the head of the jet increase with increase of stand-off distance. The penetrations by elements ΔL_{C} to ΔL_{L} increase with increase of stand-off distance until a maximum value is reached at the 'optimum stand-off distance'. The magnitude of optimum stand-off distance decreases from ΔL_{C} to ΔL_{L} elements of the jet. As stand-off distance increases beyond the optimum, the penetration by each element decreases at a rate depending upon the break-up factor, the waver and the target strength correction factor. The penetration by individual elements increases from the head to the tail end of the jet at small stand-off distances. The total penetration is the summation of penetrations by all jet elements. The portion of the jet that is sufficiently well aligned to contribute to penetration becomes progressively smaller as stand-off distance increases.

Curves of the velocity of the impinging jet element and velocity of penetration as functions of the instantaneous depth of penetration in the target are shown in figures 4 and 5. The experimental curves are not shown, as in most of the cases the theoretical curves show a reasonably good fit with the experimental curves; except in figure 5, the values of U up to 2 cm penetration at large stand-off distances. This is expected as the elements at the head of the jet remain continuous even at large stand-off distances and γ for the elements is unity. Typical plots of depth of penetration as a function of time are shown in figure 6. This suggests that the penetration velocity (slope dp/dt of the curves) increases with the increase of stand-off distance until it reaches a maximum at $S = 7{\cdot}62$ cm. The slopes of the curves decrease systematically as the stand-off distance increases beyond 7·62 cm.

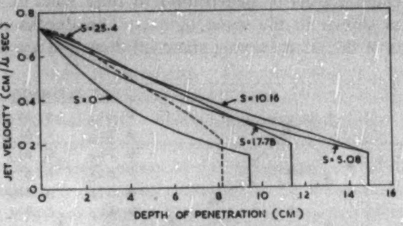


Figure 4. Jet velocity as a function of depth of penetration in mild steel for several representative stand-off distances.

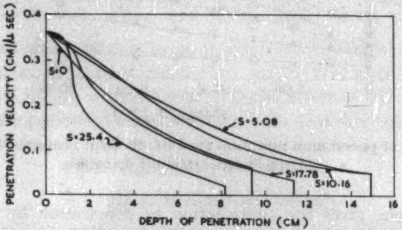


Figure 5. Velocity of penetration as a function of depth of penetration in mild steel for several representative stand-off distances.

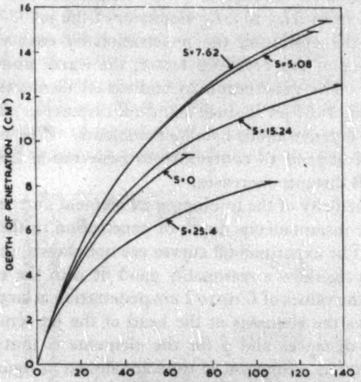


Figure 6. Depth of penetration into mild steel as a function of time for several representative stand-off distances. The instant of impact of the front of the jet on the face of the target is referred to as zero time.

The ratio V/U is evaluated for each jet element individually as a function of the distance the element has travelled. A plot of V/U against Z is shown in figure 7, where $Z(=S_B+S_D)$ is the axial distance from the top of the cone to the point of impingement of a jet element. The points shown on the plot represent values of the data at all eleven stand-off distances. At relatively small Z, $V/U=2$. As Z increases, the scatter of individual values of V/U also increases. At large stand-off distances we are not justified in combining the values of V/U for different elements of the jets in a single plot, although at relatively small stand-off distances all the jet elements can be fitted by the same relation. The experimental curve is also shown in figure 7. The experimental values of V/U at larger Z also show quite a good deal of scatter. The quantitative agreement seems to be good between the theoretical and experimental values.

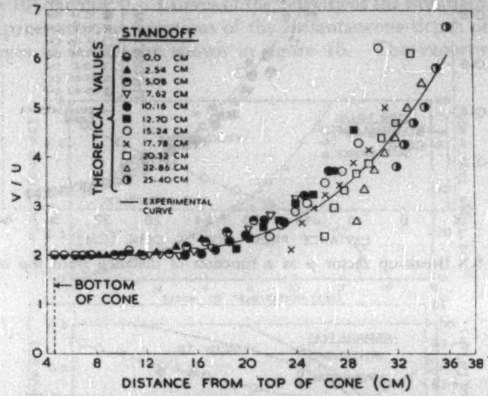


Figure 7. Ratio of jet velocity to penetration velocity in mild steel as a function of distance from top of the cone.

The break-up factor γ is unity as long as an element of the jet is a continuous one. The elements ΔL_A and ΔL_B at the head of the jet remain continuous up to the maximum stand-off distance. The calculated S_B for different elements ΔL_C, ΔL_D, ΔL_E, ΔL_F and ΔL_G of the jet are 22·2 cm, 17·4 cm, 15·8 cm, 14·3 cm and 14·2 cm respectively; and S_B for elements ΔL_H to ΔL_L is approximately 12 cm. Figure 8 shows the values of γ for each jet element and at all eleven stand-off distances as a function of Z. There is a considerable scatter in the values of γ at about 12 cm to 18 cm. This is expected as the values of S_B are different for different elements of the jet. It seems that all jet elements ΔL_C to ΔL_L can be fitted with the same relation. Eichelberger (1956) assumed $\sigma=0$ for mild steel targets and calculated γ by using data from $(V/U, Z)$ curves and these values of γ are also shown in figure 8. The effect of assigning a finite value to σ is to slightly raise the values of γ, especially at large Z; and this perhaps explains Eichelberger's lower values of γ at large Z than the mean theoretical values.

The target strength correction factor $1-k\sigma$ is plotted as a function of the depth of penetration at several representative stand-off distances in figure 9. It seems

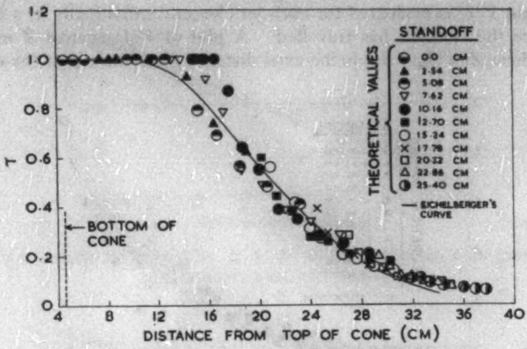


Figure 8. Break-up factor γ as a function of distance from top of cone.

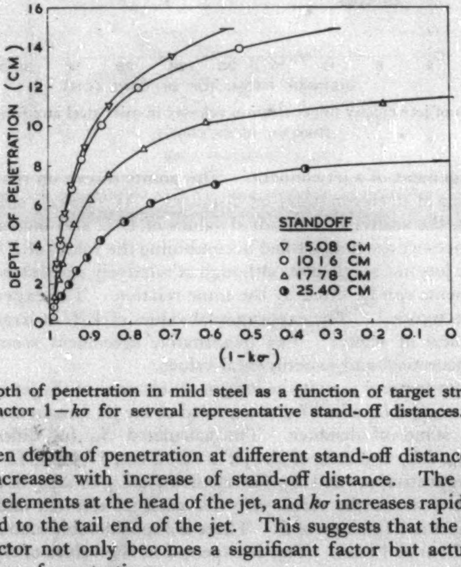


Figure 9. Depth of penetration in mild steel as a function of target strength correction factor $1-k\sigma$ for several representative stand-off distances.

that for a given depth of penetration at different stand-off distances, the value of $k\sigma$ rapidly increases with increase of stand-off distance. The value of $k\sigma$ is negligible for elements at the head of the jet, and $k\sigma$ increases rapidly for elements from the head to the tail end of the jet. This suggests that the target strength correction factor not only becomes a significant factor but actually dominates the latter stages of penetration.

3.2. *Effects of Varying Target Materials*

Similar calculations are made at constant stand-off distance 10·16 cm with targets of hardened alloy steel, 24 ST aluminium and lead. The densities of alloy steel, aluminium and lead are taken as 7·8, 2·7 and 11·37 g cm^{-3} respectively. The alloy steel has a yield strength of $11{\cdot}6\times10^9$ dyn cm^{-2}, the aluminium $2{\cdot}8\times10^9$ dyn cm^{-2}, and lead has negligible strength. For the alloy steel the experimental curves of V_{j} and t as functions of P agree with the theoretical curves for σ_{t} equal to twice the static yield strength of the target material; for the 24 ST aluminium, better agreement is obtained by using $\sigma_{\mathrm{t}} = 19{\cdot}6$ dyn cm^{-2}, a value equal to seven times the static yield strength, and for lead σ_{t} is taken as zero (i.e. $\sigma = -2 \times 10^{9}$ dyn cm^{-2}). Curves of the velocity of the impinging jet element and time of penetration as functions of the instantaneous depth of penetration in several types of targets are shown in figure 10. The experimental values are also shown in the figures. The deep penetration in lead is due to the contribution of the relatively very slow particles (i.e. $V_{\mathrm{J}} < 1500$ m sec^{-1}) at the tail end of the jet (Eichelberger 1956). They have too little energy to cause deformation in a target of any considerable strength and merely accumulate at the bottom of the hole.

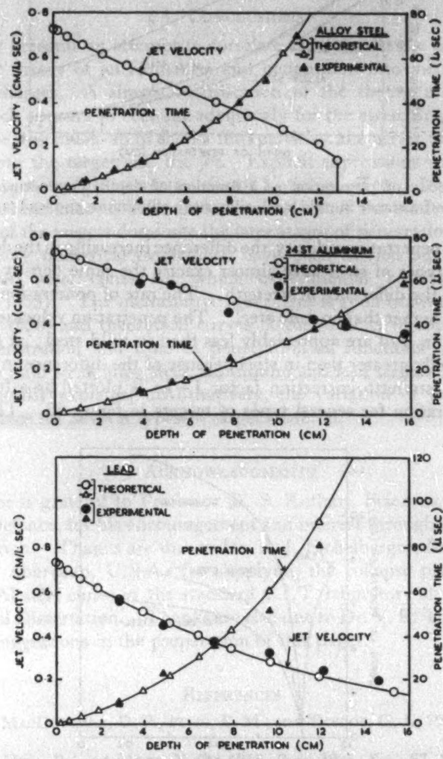


Figure 10. Jet velocity and time of penetration as functions of depth of penetration at 10·16 cm stand-off distance in alloy steel, aluminium and lead targets.

The penetration velocities as functions of depths of penetration in different targets are shown in figure 11. The rate of penetration in alloy steel is initially not significantly different from that in mild steel; thereafter the alloy steel plate yields a smaller penetration velocity, the difference increasing as the depth increases. Since the two types of steel have almost exactly the same density the difference must be due to the difference in strength. The rate of penetration in aluminium is considerably higher than in mild steel. The penetration velocities by the faster parts of the jet in lead are appreciably less than in mild steel; at greater depths the rate is actually greater than in steel because of the difference in strength.

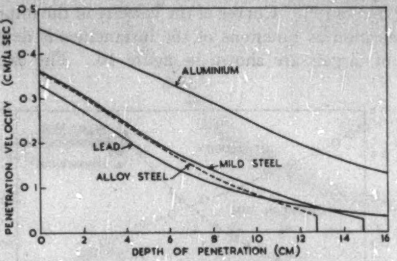


Figure 11. Velocity of penetration as a function of depth of penetration at 10·16 cm stand-off distance in mild steel, alloy steel, aluminium and lead targets.

The target strength correction factor $1-k\sigma$ is plotted as a function of the depth of penetration for several types of targets in figure 12. The value of $k\sigma$ is negligible for elements at the head of the jet, and $k\sigma$ increases rapidly for elements from the head to the tail end of the jet. At a given depth of penetration, the value of $k\sigma$ in alloy steel is considerably higher and in aluminium appreciably less than in mild steel. In lead $\sigma<0$, so $1-k\sigma$ becomes more than unity.

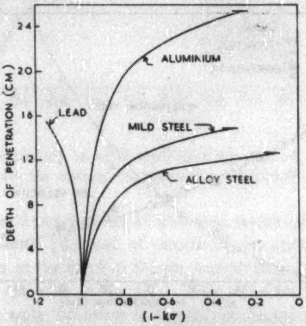


Figure 12. Depth of penetration at 10·16 cm stand-off distance as a function of target strength correction factor $1-k\sigma$ in mild steel, alloy steel, aluminium and lead targets.

The velocity of penetration of water (taking $\sigma=0$ and $\rho_t=1$ g cm^{-3}) by the elements at the head of the jet is about 5000 m sec^{-1}. It is interesting to note that experimental value of the velocity of penetration of water by metallic jets is also about 5000 m sec^{-1} (Pugh *et al.* 1951).

Considered as a whole, the good fit of the theoretical and experimental curves on various target materials add further evidence that, with this simple modification, the theory provides a surprisingly accurate description of the penetration process.

§ 4. CONCLUSIONS

This paper presents an attempt to correlate theoretically the non-steady state hydrodynamic theory of jet formation and penetration into metallic targets by lined cavity charges. A simple modification of the theory of penetration is described which appears to account adequately for the stretching of the jet as it travels in space, the break-up of the jet into particles, the waver of the jet and the strength of both the target and the jet. Explicit expressions for the break-up factor γ and target strength correction factor $1-k\sigma$ are given. It appears that the targets offer a resistance to deformation several times the static yield strength. The strength of the targets dominate the later stages of penetration. In order to check the theory numerical evaluations have been carried out in the case of standard M9Al steel cones for which experimental results are available (Eichelberger 1955). The quantitative agreement is good between experimental (Eichelberger 1956) and theoretical curves of the velocity of the impinging jet, velocity of penetration and time of penetration as functions of the depth of penetration in mild steel, alloy steel, aluminium and lead targets. The modified theory successfully explains, quantitatively, the variation of penetration with stand-off distance for several types of target material.

ACKNOWLEDGMENTS

The author is grateful to Professor D. S. Kothari, Scientific Adviser to the Ministry of Defence, for his encouragement and interest throughout the course of this investigation. Thanks are due to Dr. R. J. Eichelberger, Ballistic Research Laboratories, Aberdeen, U.S.A., for supplying the collapse parameters of the standard M9Al steel cones in the standard C.I.T. laboratory charge and a copy of his Doctoral Dissertation. Thanks are also due to Dr. V. R. Thiruvenkatachar for valuable suggestions in the preparation of this paper.

REFERENCES

BIRKHOFF, G., MACDOUGALL, D. P., PUGH, E. M., and TAYLOR, G. I., 1948, *J. Appl. Phys.* **19**, 563.

BISHOP, R. F., HILL, R., and MOTT, N. F., 1945, *Proc. Phys. Soc.*, **57**, 147.

EICHELBERGER, R. J., 1955, *J. Appl. Phys*, **26**, 398 ; 1956, *Ibid.*, **27**, 63.

EICHELBERGER, R. J., and PUGH, E. M., 1952, *J. Appl. Phys.*, **23**, 537.

EVANS, W. M., and PACK, D. C., 1951, *Proc. Phys. Soc.* B, **64**, 303.

KOLSKY, H., SNOW, C. I., and SHEARMAN, A. C., 1949, *Research*, **2**, 92.

PACK, D. C., and EVANS, W. M., 1951, *Proc. Phys. Soc.* B, **64**, 298.

PUGH, E. M., VON HEINE-GELDERN, R. FONER, S., and MUTSCHLER, E. C., 1951, *J. Appl. Phys.*, **22**, 487.

SINGH, S., 1953, *Proc. Nat. Inst. Sci. India*, **19**, 583; 1954, *Defence Sci. J. India*, **4**, 125.

RESEARCH NOTES

The Splittings induced in Levels by Crystal Fields of High Symmetry

By B. R. JUDD

The Clarendon Laboratory, University of Oxford

MS. received 6*th May* 1957, *and in final form* 11*th June* 1957

An ion situated in a crystal lattice is subject to the electric field produced by the surrounding charge system. The effect of this field can be represented by including an expression of the form

$$V = \sum_{n > m} A_n{}^m V_n{}^m$$

in the Hamiltonian (Bleaney and Stevens 1953). The $V_n{}^m$ are potential functions that transform like the corresponding spherical harmonics, $Y_n{}^m$; the constants $A_n{}^m$ depend on the lattice. The splitting of a level characterized by a value of J, the quantum number of the total angular momentum, may be found in the usual way by finding the matrix elements of V in JJ_z quantization and then constructing and solving the secular determinant. When the crystal field possesses a high degree of symmetry, it is usually found that many roots are repeated. In one sense this is an entirely expected result, since the degeneracies of the components of a level can be easily found by group theory (Bethe 1929), and the characters of the class consisting of the identity element (which determine the degeneracies) generally increase with the symmetry exhibited by the crystal field. However, in the process of finding the roots of the secular determinant, there is often no indication that some of the roots will be found to coincide; for this reason, and also because of the repetition that is incurred, this approach is rather unsatisfactory. Moreover, it is difficult to avoid the feeling that a more direct method exists, especially since many of the roots of what appear to be complex determinantal equations turn out to be very simple. A description of such a method forms the subject of this note.

To make the procedure definite, consider the splitting induced in a $J = 4$ level by a cubic field. Kynch (1937) has examined this level, among others, using the method outlined above. As Bethe (1929) showed, such a level breaks up into four components, corresponding to the irreducible representations Γ_1, Γ_3, Γ_4 and Γ_5 of the cubic group O. The conventional method for finding the positions of these components is to choose a fourfold axis as the z-axis and to construct the secular determinant. This factorizes into four smaller determinants, corresponding to the J_z combinations $(4, 0, -4)$, $(3, -1)$, $(2, -2)$ and $(1, -3)$. Now if one could tell, without actually solving the determinantal equations, how the nine roots distribute themselves among the four Γ_i, then it would be possible to write down a number of equations expressing the fact that the sum of the roots of a determinantal equation is equal to the sum of the diagonal elements. Detailed calculations reveal, for example, that the three roots corresponding to J_z admixtures of the type $(4, 0, -4)$ give the energies of the components Γ_1, Γ_3 and Γ_4. Thus, if E_i is the energy of Γ_i,

$$E_1+E_3+E_4=\langle 4|V|4\rangle+\langle 0|V|0\rangle+\langle -4|V|-4\rangle.$$

If a sufficient number of equations of this sort could be found, then clearly one would have a method for finding the E_i. Values of $\langle J_z|V|J_z\rangle$ are easily calculated in terms of parameters of the form $A_n^m\langle r^n\rangle$ by the methods of Stevens (1952).

Fortunately, it is not difficult to determine the type of J_z combinations that characterize the eigenfunctions spanning the Γ_i. In general, the eigenfunctions corresponding to the J_z admixtures $(\ldots 4, 0, -4, \ldots)$, $(\ldots 3, -1, -5, \ldots)$, $(\ldots 2, -2, -6, \ldots)$ and $(\ldots 1, -3, -7, \ldots)$ transform under rotations of 0, $\pi/2$, π and $3\pi/2$ about the z-axis in exactly the same way as the four irreducible representations of the group C_4; hence the required J_z combinations may be found immediately by performing the reduction $O\to C_4$, that is, by expressing the irreducible representations Γ_i of O in terms of the irreducible representations of C_4. It is convenient to label the latter by the smallest non-negative value of J_z in the corresponding sequence of J_z values. The reduction $O\to C_4$ is given in the table.

The Reductions $O\to C_4$, C_3, C_2

	$O\to C_4$	$O\to C_3$	$O\to C_2$
Γ_1	(0)	(0)	(0)
Γ_2	(2)	(0)	(1)
Γ_3	(0) (2)	(1) (2)	(0) (1)
Γ_4	(0) (1) (3)	(0) (1) (2)	$(0)\ (1)^2$
Γ_5	(1) (2) (3)	(0) (1) (2)	$(0)^2\ (1)$
Γ_6	$(\tfrac{1}{2})\ (\tfrac{7}{2})$	$(\tfrac{1}{2})\ (\tfrac{5}{2})$	$(\tfrac{1}{2})\ (\tfrac{3}{2})$
Γ_7	$(\tfrac{3}{2})\ (\tfrac{5}{2})$	$(\tfrac{1}{2})\ (\tfrac{5}{2})$	$(\tfrac{1}{2})\ (\tfrac{3}{2})$
Γ_8	$(\tfrac{1}{2})\ (\tfrac{3}{2})\ (\tfrac{5}{2})\ (\tfrac{7}{2})$	$(\tfrac{1}{2})\ (\tfrac{3}{2})^2\ (\tfrac{5}{2})$	$(\tfrac{1}{2})^2\ (\tfrac{3}{2})^2$

It can be seen that (2) occurs in the reduction of Γ_2, Γ_3 and Γ_5, and hence for the case of $J=4$ considered above, in which Γ_2 does not occur,

$$E_3+E_5=\langle 2|V|2\rangle+\langle -2|V|-2\rangle.$$

The equations obtained by considering (1) and (3) turn out to be the same, a result of the fact that (1) and (3) are conjugate representations of C_4. Collecting the three equations together, and working out the matrix elements of V for a configuration of the form f^n,

$$\begin{aligned} E_1+E_3+E_4&=46a-12b,\\ E_3+E_5&=-22a+44b,\\ E_4+E_5&=-12a-16b, \end{aligned}$$

in which

$$a=60\beta A_4^0\langle r^4\rangle \quad\text{and}\quad b=1260\gamma A_6^0\langle r^6\rangle$$

where β and γ are the 'operator equivalent' factors for the level in question, as defined by Stevens (1952). This method is similar to that used by Bethe (1929), and for small values of J sufficient equations can be found to solve for the E_i. For $J=4$ a further equation is required. Bethe was able to complete his analysis of this level by making use of the eigenfunctions he had previously found; this can be avoided, however, by considering the reduction of O not only to C_4, but also to other sub-groups of O. For example, taking a three-fold axis for the

z-axis, and noting the occurrence of the irreducible representation (0) of C_3 as given in the table,

$$E_1+E_4+E_5=-24a'-54b',$$

where $a'=60\beta B_4^0\langle r^4\rangle$ and $b'=1260\gamma B_6^0\langle r^6\rangle$. B_4^0 and B_6^0 are the coefficients of V_4^0 and V_6^0 in V when the three-fold axis is chosen as the z-axis; they are related to the A_n^m through the equations

$$B_4^0=-\frac{2}{3}A_4^0,\qquad B_6^0=\frac{16}{9}A_6^0.$$

There are now sufficient equations to solve for the E_i, and it is found that

$$\begin{array}{ll} E_1=28a-80b, & E_4=14a+4b, \\ E_3=4a+64b, & E_5=-26a-20b. \end{array}$$

These results agree with those found by Kynch (1937). When required, a further equation, which is linearly independent of the ones already obtained, may be found by considering $O\rightarrow C_2$. The z-axis must be so chosen that C_2 is not contained within C_4, otherwise no new information will be gained. The coefficients of V_4^0 and V_6^0 in V are now $-\frac{1}{4}A_4^0$ and $-(13/8)A_6^0$ respectively.

When two or more components of a level correspond to the same irreducible representation Γ_i, then this method will only give their mean energy. This is a drawback, of course, but detailed calculations show that the expressions for the E_i belonging to Γ_i that occur more than once are generally complicated functions of the crystal field parameters, and it seems most unlikely that simple group theory arguments could circumvent the calculations.

Recently, Hayes (unpublished) has observed a paramagnetic resonance in crystals of calcium fluoride containing a small amount of erbium. In some cases, the erbium ion, whose ground level possesses a value of 15/2 for J, is subject to a crystal field of cubic symmetry. The problem of the splitting of the ground level illustrates well the power of the method that has been developed.

Following the conventional approach, the secular equations for the energies of the components of the ground level are

$$\begin{array}{c|cccc|l} J_z & \pm 13/2 & \pm 5/2 & \mp 3/2 & \mp 11/2 & \\ \pm 13/2 & -91a-117b-E & (a-3b)\sqrt{5005} & 0 & 0 & \\ \pm 5/2 & (a-3b)\sqrt{5005} & 23a+45b-E & 42(a+b)\sqrt{15} & 0 & =0, \\ \mp 3/2 & 0 & 42(a+b)\sqrt{15} & 129a-25b-E & (5a-7b)\sqrt{429} & \\ \mp 11/2 & 0 & 0 & (5a-7b)\sqrt{429} & -221a-39b-E & \end{array}$$

where

$$a=60\beta({}^4I_{15/2})A_4^0\langle r^4\rangle \qquad \text{and} \qquad b=13\,860\gamma({}^4I_{15/2})A_6^0\langle r^6\rangle,$$

and another equation of comparable complexity for

$$J_z=(\pm 15/2,\ \pm 7/2,\ \mp 1/2,\ \mp 9/2).$$

The off-diagonal elements may be found by means of the operator equivalences

$$V_4^4=\sum(x^4-6x^2y^2+y^4)\equiv\tfrac{1}{2}\beta\langle r^4\rangle[J_+^4+J_-^4],$$

$$\begin{aligned} V_6^4=\sum(11z^2-r^2)(x^4-6x^2y^2+y^4)\equiv\tfrac{1}{4}\gamma\langle r^6\rangle[&(11J_z^2-J(J+1)-38)(J_+^4+J_-^4) \\ &+(J_+^4+J_-^4)(11J_z^2-J(J+1)-38)], \end{aligned}$$

or by using the formulae given by Kynch (1937).

It is apparent that the solution of these determinantal equations is a lengthy problem. From group theory, it can be shown that a $J=15/2$ level breaks up into $\Gamma_6+\Gamma_7+3\Gamma_8$ under the action of a cubic field. Using the table and those tables given by Stevens (1952), the equations for the E_i are rapidly found to be

$$E_6+3E_8=160a+136b,\ \ E_7+3E_8=-160a-136b\ \ \text{and}\ \ 6E_8=-268a+352b,$$

where E_8 stands for the mean energy of the three Γ_8 components. Thus

$$E_6=294a-40b,\ \ E_7=-26a-312b,\ \ E_8=\tfrac{1}{3}(-134a+176b),$$

again agreeing with the results of Kynch (1937). The erbium ion is situated at the centre of a cube of negative charges, and a few calculations reveal that $a<0$, $b>0$, and also that $a\ll b$. Thus, unless the three Γ_8 components are separated by large amounts, the component Γ_7 will be lowest. As E_7 is known, it can be substituted in the secular determinant to give the eigenstates in the usual way; a further calculation predicts an isotropic *g*-value of $34/5=6{\cdot}80$ for the doublet Γ_7, in excellent agreement with the figure of 6·78 found by Hayes.

The procedure can be applied equally well to crystal fields other than those possessing cubic symmetry. Attention has recently been paid to a field approximating to icosahedral symmetry (Judd 1957), and the theory can be easily extended to deal with this case. There is little point in using the theory when the crystal field possess only a small degree of symmetry, since irreducible representations which occur more than once are in these cases a commonplace, and little interest attaches to knowing the mean energies of component of the same kind.

REFERENCES

BETHE, H., 1929, *Ann. Phys., Lpz.*, **3**, 133.
BLEANEY, B., and STEVENS, K. W. H., 1953, *Rep. Progr. Phys.*, **16**, 108 (London : Physical Society).
JUDD, B. R., 1957, *Proc. Roy. Soc.* A, **241**, 122.
KYNCH, G. J., 1937, *Trans. Faraday Soc.*, **33**, 1402.
STEVENS, K. W. H., 1952, *Proc. Phys. Soc.* A, **65**, 209.

On the Absorption by Free Carriers in Semiconductors

BY B. DONOVAN † AND N. H. MARCH ‡
† Department of Physics, Bedford College, London
‡ Department of Physics, University of Sheffield

MS. received 23*rd May* 1957

IN a recent paper (Donovan and March 1956, to be referred to as DM) the properties associated with free carriers under the influence of alternating fields were discussed for non-degenerate semiconductors with spherical energy surfaces. The method used was essentially to average the well-known Drude–Zener expressions over the equilibrium distribution of energy, taking into account the dependence of the relaxation time τ upon energy for various scattering mechanisms. While this procedure should give good results in the lower frequency range, a more careful consideration of electron–lattice interaction is required when $\hbar\omega$ becomes comparable with the average electron energy $\boldsymbol{k}T$. Such a situation is encountered in the infra-red and it was, in fact, in connection with optical absorption that a more exact treatment (of the effect of lattice vibrations) was first described briefly by Fan and Becker (1951), based on a second order time-dependent perturbation calculation due to Fröhlich. The problem has since been discussed more fully by Schmidt (1954) and by Fan, Spitzer and Collins (1956, to be referred to as FSC) and a detailed review has been given by Fan (1956).

The purpose of this note is to attempt to assess the range of validity of the various treatments referred to above by examining a case of practical interest, namely the infra-red absorption in n-type germanium, which has been considered both experimentally and theoretically by FSC. For this purpose we first deal with the frequency dependence of the conductivity at the three temperatures 78°, 293° and 450°K considered by FSC; figure 1 shows the behaviour of σ_r/σ_0, where σ_r is the real part of the complex conductivity and σ_0 is the d.c. conductivity. In calculating these curves the relevant data for germanium have been taken from FSC, on the basis of lattice scattering only. In particular, with the effective mass $m^* = 0{\cdot}1m$, the relaxation time is given by

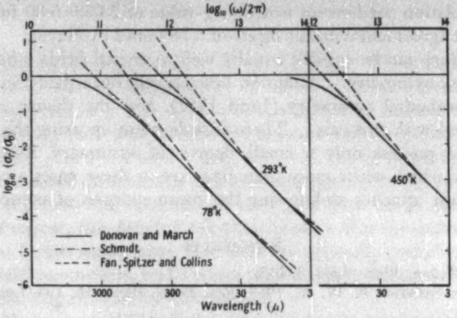


Figure 1. Variation of conductivity with frequency for n-type germanium. The frequency scale for the 450°K curves has been displaced for clarity.

$$\tau = 2{\cdot}78 \times 10^{-9} T^{-1{\cdot}66} \text{ sec.}$$

It will be seen that there is a region in which the results of DM are indistinguishable from the quantum-mechanical treatments of Schmidt and FSC, but at sufficiently high frequencies the DM curve begins to depart from the more accurate quantum-mechanical results. As a consequence of the inability of the quantum-mechanical expressions to tend to the correct limit as ω tends to 0, the results of Schmidt and FSC begin to break down below a certain frequency and this frequency increases with increasing temperature.

It is worth while now to consider briefly the re-interpretation of the experimental results for the absorption coefficient given by FSC at the highest temperature, 450°K, which they considered. For this temperature examination of figure 1 seems to indicate that at the lower frequency end the results of DM for the conductivity are preferable to the quantum-mechanical treatment in the form used by FSC. The absorption coefficient α is given by

$$\alpha = 4\pi\sigma_r/\eta c,$$

where η is the refractive index, and in figure 2 we have plotted α/n, where n is the carrier concentration, as a function of frequency (cf. FSC, figure 2). With a slight change in the effective mass, from $0{\cdot}1m$ to $0{\cdot}092m$, a very good fit with the

experimental results of FSC is obtained and the agreement is certainly appreciably better than that given by the theoretical curve of FSC.

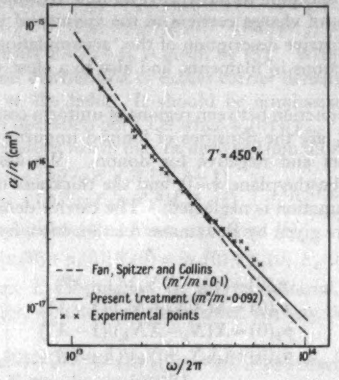


Figure 2. Variation of absorption coefficient with frequency for n-type germanium.

The conclusion which would appear to emerge from the results of this note is that, from a practical point of view, the frequency dependence of the conductivity, and hence of the absorption coefficient, is now known for free carriers with reasonable accuracy over a wide frequency range, by suitable combination of the DM and quantum-mechanical treatments. It should, however, be borne in mind that eventually more realistic ellipsoidal energy surfaces might have to be considered for germanium, to enable a closer quantitative comparison with experiment to be made.

REFERENCES

DONOVAN, B., and MARCH, N. H., 1956, *Proc. Phys. Soc.* B, **69**, 528.
FAN, H. Y., 1956, *Rep. Progr. Phys.*, **19**, 107. (London : Physical Society).
FAN, H. Y., and BECKER, M., 1951, *Semi-conducting Materials*, ed. H. K. Henisch (London : Butterworths Scientific Publications), p. 132.
FAN, H. Y., SPITZER, W., and COLLINS, R. J., 1956, *Phys. Rev.*, **101**, 566.
SCHMIDT, H., 1954, Z. *Phys.*, **139**, 433.

Low–High Conductivity Junctions in Semiconductors

BY L. W. DAVIES

Commonwealth Scientific and Industrial Research Organisation,
Radiophysics Laboratory, Sydney, Australia

MS. received 28*th May* 1957

MANY investigations have been made of the properties of junctions between regions of differing conductivity type (p–n junctions) in semiconductors such as germanium and silicon. More recently, the non-equilibrium distribution of charge carriers near junctions between regions of the same conductivity type, but of differing conductivity, has proved of interest (Low 1955, Arthur, Gibson and Gunn 1956 a, b). An analysis is presented here for the latter case, referred to as an L–H junction; L and H refer to light and heavy doping of donor or acceptor impurity. The new treatment, which takes into account the diffusion of charge carriers in the vicinity of the junction, leads in particular to a quantitative description of the 'accumulation' of minority charge carriers at L–H junctions in filaments, and also to a close understanding of the current gain.

We consider the junction between regions of uniform conductivity, designated j and k; N_j and N_k are the densities of ionized impurity centres, taken to be positive for acceptors and negative for donors. We assume $N_j > N_k$. The junction is taken to be the plane $x=0$, and the thickness of the depletion layer associated with the junction is neglected. The carrier densities at either side of the depletion layer are given by Boltzmann relationships, from which are derived (Fletcher 1957)

$$\left.\begin{aligned} p_j(0) &= (N_j - XN_k)/(1-X^2) \\ n_j(0) &= X(XN_j - N_k)/(1-X^2) \\ p_k(0) &= X(N_j - XN_k)/(1-X^2) \\ n_k(0) &= (XN_j - N_k)/(1-X^2) \end{aligned}\right\} \quad \ldots\ldots(1)$$

with

$$X = \exp\{-q(\psi - V)/\boldsymbol{k}T\}, \quad \ldots\ldots(2)$$

where, for L–H junctions, $N_k/N_j \leqslant X \leqslant 1$, ψ is the contact potential difference at the junction in equilibrium, and V is the potential difference applied to the junction.

Associated with the potential difference V there is a current through the junction, density $I = I_{\mathrm{n}} + I_{\mathrm{p}}$, and an electric field in the material to either side of the junction (E_j, E_k). If we make the approximation that the excess density of minority carriers is much less than that of the majority carriers, then the electric field may be assumed constant in the steady-state continuity equation, from which we may derive the non-equilibrium distribution of charge carriers:

$$\left.\begin{aligned} n_j(x) - n_{j0} = p_j(x) - p_{j0} &= \{n_j(0) - n_{j0}\}\exp(x/\mathscr{L}_j), & x<0 \\ n_k(x) - n_{k0} = p_k(x) - p_{k0} &= \{n_k(0) - n_{k0}\}\exp(x/\mathscr{L}_k), & x>0 \end{aligned}\right\} \quad \ldots\ldots(3)$$

where, for N_j, $N_k > 0$,

$$\mathscr{L}_{j,k}^{-1} = L_{j,k}^{-1}\left[-\left(\frac{qE_{j,k}L_{j,k}}{2\boldsymbol{k}T}\right) \pm \left\{1 + \left(\frac{qE_{j,k}L_{j,k}}{2\boldsymbol{k}T}\right)^2\right\}^{1/2}\right], \quad x \lesseqgtr 0$$

and for N_j, $N_k < 0$,

$$\mathscr{L}_{j,k}^{-1} = L_{j,k}^{-1}\left[\left(\frac{qE_{j,k}L_{j,k}}{2\boldsymbol{k}T}\right) \pm \left\{1 + \left(\frac{qE_{j,k}L_{j,k}}{2\boldsymbol{k}T}\right)^2\right\}^{1/2}\right]. \quad x \lesseqgtr 0 \quad \ldots\ldots(4)$$

In these equations n_{j0}, p_{j0}; n_{k0}, p_{k0} are the equilibrium charge carrier densities, and L_j, L_k the diffusion lengths for minority carriers.

The electron and hole components of the current density on either side of the depletion layer are given by

$$\left.\begin{aligned} I_{\mathrm{n}j}(0) &= q\mu_{\mathrm{n}}[n_j(0)E_j(0) + (\boldsymbol{k}T/q\mathscr{L}_j)\{n_j(0) - n_{j0}\}] \\ I_{\mathrm{p}j}(0) &= q\mu_{\mathrm{p}}[p_j(0)E_j(0) - (\boldsymbol{k}T/q\mathscr{L}_j)\{p_j(0) - p_{j0}\}] \\ I_{\mathrm{n}k}(0) &= q\mu_{\mathrm{n}}[n_k(0)E_k(0) + (\boldsymbol{k}T/q\mathscr{L}_k)\{n_k(0) - n_{k0}\}] \\ I_{\mathrm{p}k}(0) &= q\mu_{\mathrm{p}}[p_k(0)E_k(0) - (\boldsymbol{k}T/q\mathscr{L}_k)\{p_k(0) - p_{k0}\}]. \end{aligned}\right\} \quad \ldots\ldots(5)$$

If there is no recombination in the depletion layer we obtain simultaneous equations for $E_j(0)$, $E_k(0)$ from the conditions for continuity of I_{n}, I_{p} at the junction:

$$\left.\begin{aligned} n_j(0)E_j(0)+(\boldsymbol{k}T/q\mathscr{L}_j)\{n_j(0)-n_{j0}\}&=n_k(0)E_k(0)+(\boldsymbol{k}T/q\mathscr{L}_k)\{n_k(0)-n_{k0}\} \\ p_j(0)E_j(0)-(\boldsymbol{k}T/q\mathscr{L}_j)\{p_j(0)-p_{j0}\}&=p_k(0)E_k(0)-(\boldsymbol{k}T/q\mathscr{L}_k)\{p_k(0)-p_{k0}\} \end{aligned}\right\} \ldots(6)$$

The total current density $I=I_{\mathrm{n}}+I_{\mathrm{p}}$ is given by (5). With the aid of equations (1)–(6) the current and distribution of excess carrier density may be determined analytically, as well as the field. It should be emphasized that these results are valid for low-level operation only, i.e. when $n_k(0)\ll p_{k0}$.

Some useful expressions which may be derived from those given above include

$$\left.\begin{aligned} E_j&=\pm(\boldsymbol{k}T/q\mathscr{L}_j)(\mathscr{L}_j^{2}/L_j^{2}-1) \\ E_k&=\pm(\boldsymbol{k}T/q\mathscr{L}_k)(\mathscr{L}_k^{2}/L_k^{2}-1) \end{aligned}\right\} N_j,\ N_k \gtreqless 0, \qquad \ldots\ldots(7)$$

both derived from equations (4). From equations (6),

$$\{n_j(0)+p_j(0)\}E_j(0)=\{n_k(0)+p_k(0)\}E_k(0). \qquad \ldots\ldots(8)$$

In the case of an L–H junction in p-type material, for example, when $X\ll\mathscr{L}_j/(-\mathscr{L}_k)$ or unity, whichever is the smaller, we may derive the approximations

$$\begin{aligned} E_k(0)&\simeq(\boldsymbol{k}T/qL_k)\{n_k(0)-n_{k0}\}/\{n_k(0)\,.\,n_{k0}\}^{1/2} \\ \mathscr{L}_k&\simeq-\{n_{k0}/n_k(0)\}^{1/2}L_k \\ \mathscr{L}_j&\simeq L_j. \end{aligned}$$

From equations (1) we have

$$\left.\begin{aligned} n_j(0)-n_{j0}=p_j(0)-p_{j0}&=(1-X^2)^{-1}(X-X_0)(Xp_{j0}+n_{k0}) \\ n_k(0)-n_{k0}=p_k(0)-p_{k0}&=(1-X^2)^{-1}(X-X_0)(Xn_{k0}+p_{j0}) \end{aligned}\right\} \ldots(9)$$

where

$$X_0=\exp(-q\psi/\boldsymbol{k}T)=n_{j0}/n_{k0}=p_{k0}/p_{j0}. \qquad \ldots\ldots(10)$$

It follows that there is an excess of minority carriers on both sides of the junction when $V>0$, and a deficit when $V<0$. The excess and deficit of carriers for $V\gtreqless 0$ has suffered from a wide variety of nomenclature: if we take the case of an L–H junction in p-type material, for example, then the condition described by $n_j(V>0)$ has been described as injection, $n_k(V<0)$ as exclusion and $n_k(V>0)$ as accumulation (Low 1955), and $n_j(V<0)$ as extraction or egression (Angello and Ebert 1954).

We investigate in some detail the flow of minority carriers (electrons) near such a junction biased in the 'forward' direction ($V>0$). In the H region there is an exponential distribution of excess electrons and holes, and the minority-carrier current $I_{\mathrm{n}j}(0)$ arises for the greater part from diffusion, as may be shown with the aid of equation (5). In the L region, the drift velocity $\bar{v}_{\mathrm{n}k}$ of the electrons is given by

$$-\bar{v}_{\mathrm{n}k}=\mu_{\mathrm{n}}E_k-(-D_{\mathrm{n}}/\mathscr{L}_k)(n_k-n_{k0})/n_k$$

whence by equation (7),

$$-\bar{v}_{\mathrm{n}k}=-(D_{\mathrm{n}}/\mathscr{L}_k)\{(1-\mathscr{L}_k^{2}/L_k^{2})-(n_k-n_{k0})/n_k\}. \qquad \ldots\ldots(11)$$

Thus the drift velocity of the electrons toward the junction decreases monotonically as the junction is approached, due to the increasing effect of the opposing concentration gradient.

The effective reduction in the mobility of the minority carriers gives rise to large values of the current gain at such junctions when they are incorporated in filamentary structures (see, for example, Arthur, Gibson and Gunn 1956 b). The current multiplication factor α^*, defined such that $\alpha^* - 1$ is equal to the ratio of the effective carrier mobilities at the junction, is given by

$$\alpha^* - 1 = n_k(0)/\Gamma p_k(0), \qquad \ldots\ldots(12)$$

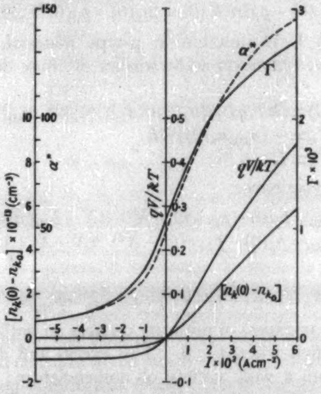
where

$$\Gamma = I_{\mathrm{n}}(0)/I \qquad \ldots\ldots(13)$$

is the proportion of the total current through the junction carried by minority carriers. From equations (5) and (7),

$$\Gamma = (-b\mathscr{L}_k/L_j)\{n_j(0) - n_{j0}\}/\{(2 - \mathscr{L}_k^2/L_k^2)p_k(0) - p_{k0}\}. \quad \ldots\ldots(14)$$

In the figure results are presented for an L–H junction in p-type germanium, calculated from the theory given above.



Calculated values of the potential drop (qV/kT), minority-carrier density $[n_k(0) - n_{k0}]$, current ratio Γ and current gain α^* as a function of current density I, for an L–H junction in p-type germanium. [Values assumed: $N_j = 10^{17}$ cm^{-3}; $N_k = 10^{14}$ cm^{-3}; $L_j = 10^{-2}$ cm; $L_k = 10^{-1}$ cm.]

A qualitative discussion of the L–H junction has been given by Arthur, Gibson and Gunn (1956 a), in which the junction is assumed to present a barrier of low permeability to the drift of minority carriers in the L region. A 'leakage velocity' $V_{0\mathrm{L}}$ is introduced, which is stated to be independent of the current I; however, according to the analysis given here, we see that, in the case of a junction in p-type material

$$V_{0\mathrm{L}} \equiv I_{\mathrm{n}k}(0)/q\{n_k(0) - n_{k0}\} = (D_{\mathrm{n}}/\mathscr{L}_j)\{(\mathscr{L}_j^2/L_j^2)n_j(0) - n_{j0}\}/\{n_k(0) - n_{k0}\}, \quad \ldots(15)$$

and hence that

$$V_{0\mathrm{L}} \simeq (XD_{\mathrm{n}}/L_j).$$

Thus it appears that the accumulation of minority carriers on the low-conductivity side of an L–H junction, biased with $V > 0$, is perhaps better described as the accompaniment to the injection of majority carriers from the H region; the gradient in concentration then leads to a decrease in the effective mobility of the minority carriers, with which is associated current gain.

ACKNOWLEDGMENTS

The author wishes to thank Dr. N. H. Fletcher for helpful discussions.

REFERENCES

ANGELLO, S. J., and EBERT, T. E., 1954, *Phys. Rev.*, **96**, 221.
ARTHUR, J. B., GIBSON, A. F., and GUNN, J. B., 1956 a, *Proc. Phys. Soc.* B, **69**, 697; 1956 b, *Ibid.*, **69**, 705.
FLETCHER, N. H., 1957, *J. Electronics*, **2**, 609.
LOW, G. G. E., 1955, *Proc. Phys. Soc.* B, **68**, 310.

Electron Microscopy of Tin Whiskers using Carbon Replicas

By D. E. BRADLEY, J. FRANKS AND P. E. RUSH
Research Laboratory, Associated Electrical Industries Ltd., Aldermaston, Berks.

MS. received 21*st April* 1957, *and in revised form* 20*th May* 1957

§ 1. INTRODUCTION

IN general, the direct examination of dense objects in the electron microscope provides little information since only a silhouette is obtained. In the case of tin whiskers, however, direct examination during growth has revealed that growth takes place at the base (Koonce and Arnold 1954, and Levy and Kammerer 1955). In addition, silhouettes have shown that many whiskers are in the form of ribbons (see figure 1, Plate I). Apart from this, silhouettes provide little additional information; no indication of the presence of surface structure is apparent, and the shape of the whisker remains obscure, except in the case of ribbons. Franks (1956) showed by means of carbon replicas (Bradley 1954 a) that some whiskers are star-shaped in section, and this method has now revealed many additional details of shape and surface structure which are described here.

§ 2. EXPERIMENTAL METHODS

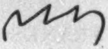
The method described by Fisher *et al.* (1954) was used to prepare whiskers. Layers of tin, 5 μ thick, electrodeposited on to 0·3 mm thick steel plates were clamped between two rigid plates. A cross section was then prepared metallographically. Compression applied to the tin layer through the clamping plates caused the whiskers to grow in about one hour.

Whiskers were removed from the clamp assembly by touching a cluster with the corner of a clean microscope slide while being viewed under a light microscope. It was found that a number of whiskers adhered to the glass. The whiskers were then coated with a layer of carbon (about 150 Å thick) by vacuum evaporation (Bradley 1954 b), and covered with a backing of the plastic Bedacryl, which was deposited from a solution in benzene. The combined film was floated on to the surface of 50% hydrochloric acid to dissolve the tin, leaving the backed replica. This was finally transferred to an electron microscope specimen support grid and the Bedacryl removed by washing in chloroform. The contrast of the replica was improved by shadowing with gold–palladium at 2 : 1.

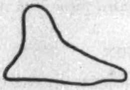
In some cases where it was thought that replicas had been distorted during processing, the whiskers themselves were shadowed before the deposition of the carbon. Here, the shadowing contrast is produced by the actual whisker rather than the replica, and any distortion of the latter can be readily detected.

§ 3. Results

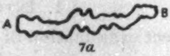
Electron micrographs of carbon replicas revealed a variety of shapes, including star-shapes, ribbons and cylinders. These forms were found to co-exist in a single growth region. A careful study of an individual replica permits an approximate profile of the visible surface to be drawn. The information required for the production of such a drawing is provided by the changes in contrast due to the shadowing material. The various forms of whisker are now considered in turn.



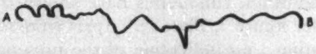
5a



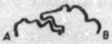
6a



7a



8a



12a

Figure 5(*a*). Diagram of a suggested section through the ribbon in figure 5.

Figure 6(*a*). Suggested section though the ribbon in figure 6.

Figure 7(*a*). Suggested section through the ribbon in figure 7 from A to B.

Figure 8(*a*). Suggested section through the ribbon in figure 8 from A to B.

Figure 12(*a*). Suggested section through the whisker shown in figure 12 from A to B.

Cylindrical Whiskers

A cylindrical whisker with little or no surface structure is shown in figure 2 (Plate I). Relatively few such cylinders were observed. Sometimes they were seen as bundles forming composite whiskers from which individual units branched (figure 3, Plate I). The single cylinders were the smallest whiskers encountered, being less than 0·25 μ in diameter.

Coarsely Ridged Whiskers

This form was also relatively rare, but was not always restricted to whiskers of small diameter. Figure 4 (Plate I) shows a replica of what is probably a whisker with a nearly perfect six-pointed star cross section. The whisker in figure 5 (Plate I) has three sharp-pointed ridges on the visible surface suggesting a profile shown in figure 5 (*a*). The shadow width indicates that the average cross section is circular.

The whisker in figure 6 (Plate I) lies somewhere between a three-pointed star and a triangle. Here, the contrast produced by the shadowing material is negligible compared with that inherent in the carbon replica, which has been tilted in the electron beam. This enables the end face of the whisker to be seen, though the view is considerably foreshortened. Its likely appearance, viewed end on, is shown in the diagram in figure 6 (*a*). This triangular whisker was unique in the selection examined, and was the only one with the end face so clearly visible.

Ribbons

This was by far the commonest form of whisker. It was at first thought that the replicas which appeared as ribbons had, in fact, collapsed from a rounder shape. However, pre-shadowed replicas and the direct examination of whiskers eliminated this possibility. The silhouette in figure 1 is due to a ribbon-shaped whisker. These ribbons are very thin, but by no means free from surface structure. Figure 7 (Plate I) shows a replica of a ribbon which is only about 1000 Å thick; this thickness can be measured directly from the point where the ribbon twists. The fact that the ribbon twists is also useful in that it permits both sides to be examined. If these are studied carefully, it can be seen that the surface structure on one side is repeated on the other, so that the cross section from A to B would appear as shown in figure 7 (*a*). The narrower ribbons, about 1 μ across, were extremely common. Wider ribbons, 4 μ to 5 μ across were also frequently found. These, as can be seen from figure 8 (Plate I), are abundantly ridged, giving a complicated cross section (A to B on figure 8) as shown in figure 8 (*a*). In addition, it was found that some of the ridges parted from the main whisker suggesting that it was made up of a number of smaller units. A number of ribbons lying flat are shown in figure 9 (Plate II). It can be seen that they are all very similar, and resemble the whisker in figure 8, which can be taken as typical of this kind.

Other Shapes

Whiskers of more complex shapes were also common, and a few examples are given here. Figure 10 (Plate II) shows a whisker with many fine ridges, some of which are of the order of 100 Å in width. The replica in figure 11 (Plate II) has tilted slightly, and can be seen to resemble a piece of celery; it could be described as a curled ribbon. Figure 12 (Plate II) shows a whisker which has split at the end. It is significant that the inner face of one of the parts is ridged, and it is reasonable to suppose that the two portions key together as shown in figure 12 (*a*), which represents a cross section from A to B. It is possible that the units of many of the composite whiskers fit together in this way. These whiskers may well originate from two separate but adjacent sources.

§ 4. CONCLUSION

Existing theories of growth mechanism do not lead one to expect the large number of different shapes of whiskers found here. It is interesting to note that no straightforward hexagonal or cubic rods, which were at first expected, have been found. The idea of composite sources is certainly supported by some of the above observations, but the mechanism of formation of the individual units may be complicated.

ACKNOWLEDGMENTS

The authors would like to thank Mr. R. Phillips and Dr. P. Charsley of this Laboratory for their advice, and also Dr. T. E. Allibone, Director of the Research Laboratory, Associated Electrical Industries Ltd., for permission to publish this paper.

REFERENCES

BRADLEY, D. E., 1954 a, *Brit. J. Appl. Phys.*, **5**, 96; 1954 b, *Ibid.*, **5**, 65.
FISHER, R. M., DARKEN, L. S., and CARROLL, K. G., 1954, *Acta Metallurg.*, **2**, 370.
FRANKS, J., 1956, *Nature, Lond.*, **177**, 984.
KOONCE, S. E., and ARNOLD, S. M., 1954, *J. Appl. Phys.*, **25**, 134.
LEVY, P. W., and KAMMERER, O. F., 1955, *J. Appl. Phys.*, **26**, 1182.

The Infra-Red Luminescence of Iodine

BY J. FRANKS†
Department of Physics, University of Hull

Communicated by G. F. J. Garlick ; *MS. received* 26*th April* 1957

§ 1. INTRODUCTION

THE infra-red luminescence of the solid halogens has been reported by Dumbleton (1955). In the present investigation a detailed study of the excitation and emission spectra of solid iodine and of their variation with temperature has been made. From the observations, the energy configurational coordinate curves have been constructed for the normal and excited states and compared with the data already available for the iodine molecule.

§ 2. EXPERIMENTAL TECHNIQUES AND OBSERVATIONS

The iodine samples were mounted in a vacuum cryostat using either crystals from a melt, or thin films evaporated on to glass slides. The excitation and emission characteristics were unaltered by the form of the specimen. To obtain excitation spectra the iodine was excited with monochromatic light of selected frequency and the total luminescence output measured using appropriate filters. For emission spectra a wider band of excitation frequencies about that for maximum excitation was selected by optical filters and the luminescence output was analysed spectrally. A Leiss type double prism monochromator and a lead sulphide photoconductive cell were used for the experiments and the radiation was modulated at 800 c/s by use of a rotating sectored disc. The cell output was amplified by a 800 c/s tuned amplifier and the rectified output of the latter fed to a pen-chart recorder.

Figure 1(*a*) shows typical excitation and emission spectra which are very similar to those obtained by Dumbleton. However, some errors in correction were found for the earlier spectra and the present curves are corrected for the variation with wavelength of source intensity, monochromator dispersion and detector sensitivity. The peak positions and half widths of the excitation and emission curves are independent of temperature. Figure 2 shows the variation of the relative luminescence efficiency η with temperature. Even at liquid air temperature the curve shows no sign of becoming horizontal with fall of temperature. Figure 1(*b*) gives a typical absorption spectrum for a thin film specimen and a diffuse reflection spectrum for a polycrystalline layer. Magnesium oxide was used as a 'flat response' reference surface for the reflection spectra. An absorption peak appears at $1{\cdot}12\,\mu$ not associated with excitation of luminescence. The absorption spectra represent a mean absorption and average the effects of the marked anisotropy of iodine (see Moss 1952).

---

† Now at the Siemens Edison Swan Research Laboratory, Brimsdown, Enfield, Middx.

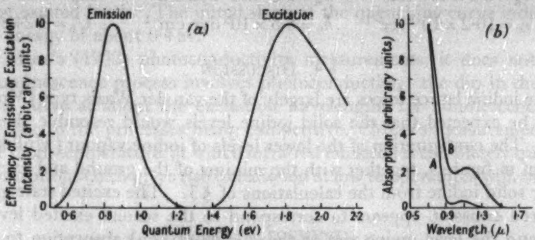


Figure 1 (*a*) Emission and excitation spectra for the luminescence of solid iodine in the temperature range -70 to -180°c. (*b*) Absorption in solid iodine. A. From optical transmission measurements at room temperature. B. From diffuse reflection measurements at room temperature and at -150°c (no significant difference in the reflection spectra at the two temperatures could be detected).

§ 3. Energy Configurational Coordinate Curves for Solid Iodine

The data for the excitation and emission spectra were used to construct the energy configurational coordinate curves for the normal and excited states by adoption of the method due to Klick and Schulman (1952). The interatomic distance in the iodine molecule is taken as the spatial parameter and to use the theory attempts were made to fit Gaussian curves to the spectra. The emission spectrum of figure 1 fits the Gaussian form quite closely but the excitation spectrum shows a Gaussian form only at energies lower than that for the peak. This part of the curve was therefore used.

The data of figure 1(*a*) give a half width $\Delta E_{\mathrm{m}}=0{\cdot}24$ ev for the emission spectrum and a half width $\Delta E_{\mathrm{x}}=0{\cdot}42$ ev for the excitation spectrum. The emission curve has a maximum at $U_{\mathrm{m}}=0{\cdot}86$ ev and that of the excitation curve is at $U_{\mathrm{x}}=1{\cdot}83$ ev. The mass of the vibrating system involved is assumed to be that of the iodine atom ($M=2{\cdot}12\times10^{-22}$ g).

The displacement of the excited state from the ground state in the energy U_0 and in the distance coordinate X_0 can be obtained from the relations

$$\nu_{\mathrm{g}}=(2\pi)^{-1}(k_{\mathrm{g}}/M)^{1/2} \qquad \nu_{\mathrm{e}}=(2\pi)^{-1}(k_{\mathrm{e}}/M)^{1/2}$$

$$U_{\mathrm{x}}=U_0+\tfrac{1}{2}k_{\mathrm{e}}X_0^2 \qquad U_{\mathrm{m}}=U_0-\tfrac{1}{2}k_{\mathrm{g}}X_0^2$$

combined with

$$\Delta E_{\mathrm{x}}=1{\cdot}355\times10^{-13}k_{\mathrm{e}}X_0(\nu_{\mathrm{g}}/k_{\mathrm{g}})^{1/2}$$

$$\Delta E_{\mathrm{m}}=1{\cdot}355\times10^{-13}k_{\mathrm{g}}X_0(\nu_{\mathrm{e}}/k_{\mathrm{e}})^{1/2}$$

where ν_e and ν_g are the vibrational frequencies of the system in the ground and excited states, and the corresponding force constants are k_e and k_g. The equations for the half widths apply to curves at 0°K, but as no variation in shape of the curves with temperature could be detected, the measured values of ΔE_x and ΔE_m were taken, giving the following results

$$\nu_g = 8{\cdot}1 \times 10^{12}\,\text{sec}^{-1} \qquad k_g = 5{\cdot}5 \times 10^5\,\text{dyn cm}^{-1} \qquad X_0 = 0{\cdot}15\,\text{Å}$$

$$\nu_e = 10{\cdot}2 \times 10^{12}\,\text{sec}^{-1} \qquad k_e = 8{\cdot}6 \times 10^5\,\text{dyn cm}^{-1} \qquad U_0 = 1{\cdot}24\,\text{ev}$$

§ 4. DISCUSSION

As the iodine lattice forces are largely of the van der Waals type (Moss 1952), it might be expected that the solid iodine levels would resemble those of the vapour. The configuration of the lower levels of iodine vapour (Mulliken 1932) are shown in figure 3, together with the minima of the ground and excited state levels for solid iodine from the calculations of § 3. The excited state giving rise to infra-red emission appears to correspond to the second excited level of the vapour, and the level giving rise to the 1·15 ev (1·12 μ) absorption to the first. No emission was observed resulting from the latter absorption, and none would be expected if the solid iodine levels here resemble those of the vapour. With this assumption, part of the first excited state has been sketched in the figure, by taking the minimum of this level to be displaced in proportion to that of the second level.

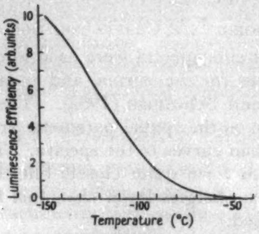


Figure 2

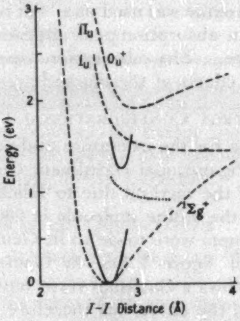


Figure 3

Figure 2. Variation of luminescence efficiency with temperature for solid iodine.

Figure 3. Energy configurational coordinate (interatomic distance in molecule) diagrams for electronic states of iodine. Broken curves: ground and excited states for molecules in the vapour phase (from Mulliken 1932); full curves: ground and excited states for solid iodine derived from data on luminescence; dotted curve: excited state giving 1.12μ absorption band in solid iodine (corresponds to 1_u state in vapour phase).

This can only be a rough approximation, however, since for the ground state, the intra-molecular distance is extended by van der Waals attraction from 2·65 Å in the vapour to 2·70 Å in the solid phase (Moss 1952), whereas the minimum of the second excited state occurs nearer the origin for the solid, indicating that repulsive forces are becoming important.

No change was detected in the shape of the emission and excitation curves for different temperatures, confirming that the vibrational frequencies are high (Klick and Schulman 1952). The emission intensity, however, varied quite rapidly with temperature (figure 2). The increase in emission with decrease in temperature gave no indication of levelling out, even as liquid air temperatures were approached; this was due perhaps to a competing radiationless transition via the first excited level. The initial slope of the quenching curve indicates an activation energy of about 0·4 ev.

From Moss's (1952) photoconductivity measurements, it does not appear that the luminescence process involves photoconduction; the dip in the photoconduction curve at about the excitation wavelengths for luminescence might in part be due to the processes being competitive. Moss's measurements were made near room temperature, at which infra-red emission is completely quenched, but the absorption which gives rise to emission at low temperatures will, of course, still occur.

ACKNOWLEDGMENT

I wish to thank Professor G. F. J. Garlick for many valuable discussions and for extending to me the facilities of his laboratory.

REFERENCES

DUMBLETON, M. J., 1955, *Proc. Phys. Soc.* B, **68**, 53.

KLICK, C. C., and SCHULMAN, J. H., 1952, *J. Opt. Soc. Amer.*, **42**, 910.

MOSS, T. S., 1952, *Photoconductivity in the Elements* (London: Butterworths Scientific Publications).

MULLIKEN, R. S., 1932, *Rev. Mod. Phys.*, **4**, 3.

LETTERS TO THE EDITOR

High Temperature Susceptibility Expansions for the Close Packed Hexagonal Lattice

Stimulated by a recent paper in the *Proceedings of the Physical Society* in which Carter and Stevens (1956) discuss the magnetic susceptibility of two polymorphic forms of manganese sulphide, one face-centred cubic and the other close-packed hexagonal, we have examined the corresponding theoretical high temperature expansions following the method of Van Vleck (1937) and Opechowski (1938).

The problem of calculating coefficients in high temperature expansions can be reduced to the evaluation of certain lattice constants at each stage (Rushbrooke and Wood 1955, Domb and Sykes 1957). The first five terms in the expansion of the susceptibility of the Heisenberg model for spin 1/2 have been calculated by Rushbrooke and Wood (1955) who express the coefficients $a_1 \ldots a_5$ in general form for any bravais lattice. Their expressions also hold for the close-packed hexagonal lattice but as in general the coefficients after a_4 need modification if the lattice is not a bravais lattice, we have calculated a_5 and a_6 as linear functions of the lattice constants p_{nx} (defined by Domb and Sykes (1957)) and find

$$\begin{aligned} a_5 = [(\sigma+1)&(120\sigma^4 - 480\sigma^3 + 640\sigma^2 + 160\sigma - 104) \\ &- 12p_3(300\sigma^2 - 640\sigma - 140) \\ &- 16p_4(200\sigma - 160) - 2000p_5 + 1440p_{5a}]/3840 \\ a_6 = [(\sigma+1)&(720\sigma^5 - 3600\sigma^4 + 6960\sigma^3 - 4064\sigma^2 - 5168\sigma - 1232) \\ &- 12p_3(2400\sigma^3 - 8280\sigma^2 + 3504\sigma - 816) \\ &- 16p_4(1800\sigma^2 - 4992\sigma) - 20p_5(1200\sigma - 1128) \qquad \ldots\ldots(1) \\ &- 14\,400p_6 + 8p_{5a}(2160\sigma + 3888) \\ &+ 9792p_{6a} + 9216p_{6b} + 29\,568p_{6c} + 3072p_{6d}]/46\,080 \end{aligned}$$

where $\sigma+1$ is the coordination number, and the a_n are those defined by Rushbrooke and Wood (1955).

Tables of the values of the constants p_{nx} for the simpler lattices are given by Domb and Sykes (1957). The relations (1) will be valid for all lattices with equivalent sites and also for finite clusters of spins if all the sites in the cluster are equivalent. The complicated reductions that lead to (1) can be checked by calculating the susceptibilities of simple clusters such as the triangle and the tetrahedron and comparing the coefficients in their expansions with those obtained from (1). The constants p_{nx} for a few such clusters are given by Domb and Sykes (1957).

The evaluation of lattice constants on the close-packed hexagonal lattice presents special difficulties. We have been able to show that the number of random walk returns to the origin after n steps is the same as for the face-centred cubic lattice and this enables the number of polygons p_n to be calculated for small n (Domb and Sykes 1957). For high order lattice constants some modifications of the simple theory are necessary and we hope to treat the matter in detail in a subsequent paper. All the constants that occur in (1) have been evaluated and the only one that differs from the corresponding constant for the face-centred cubic is p_{8a}. This has the value 36 on the face-centred cubic and 37 on the close-packed hexagonal. The corresponding values of a_0 are 2 743 803/240 (f.c.c.) and 2 743 854/240 (c.p.h.).

The difference is very small but implies a higher critical temperature for the close-packed hexagonal. We have verified that this trend is continued for higher terms of the corresponding Ising series for which the effect is even smaller, the Ising model being less structure sensitive. We infer that the close-packed hexagonal is more 'close-packed' than the face-centred cubic; it is interesting to note that a similar conclusion results from a consideration of the relative stabilities of the two lattices under van der Waals forces (see for example Barron and Domb 1954).

We are grateful to Dr. K. W. H. Stevens for drawing our attention to this problem.

King's College, London.
24th June 1957.
C. Domb.
M. F. Sykes.

Barron, T. H. K., and Domb, C., 1955, *Proc. Roy. Soc.* A, **227**, 447.
Carter, W. S., and Stevens, K. W. H., 1956, *Proc. Phys. Soc.* B, **69**, 1006.
Domb, C., and Sykes, M. F., 1957, *Phil. Mag.*, **2**, 733.
Opechowski, W., 1938, *Physica*, **4**, 186.
Rushbrooke, G. S. and Wood, P. J., 1955, *Proc. Phys. Soc.* A, **68**, 1161.
Van Vleck, J. H., 1937, *J. Chem. Phys.*, **5**, 320.

Hyperfine Structure and Nuclear Electric Quadrupole Moment of ^{17}O

The hyperfine structure of free ^{17}O atoms has been measured in a paramagnetic resonance experiment, using the apparatus described by Bowers, Kamper and Knight (1957).

The free atoms were formed by maintaining a radio-frequency discharge in a gas of molecular oxygen containing 0·75% of ^{17}O. The oxygen was supplied by the Atomic Energy Research Establishment, Harwell, in the form of water with an enhanced abundance of the heavier isotopes, and was liberated by electrolysis.

The gas was enclosed at a pressure of about ½ mm Hg in a silica tube of 1 cm internal diameter, one end of which projected into the cavity resonator. Just above the resonator the tube was surrounded by the coil of a tuned circuit, resonant at about 80 Mc/s. Power was coupled into the coil from an oscillator which was capable of delivering 100 watts at this frequency. Under appropriate tuning conditions the visible discharge can be made to extend down the tube right into the resonator. When it does so, the presence of ions in it does not lower the Q of the cavity appreciably, except at the cyclotron resonance frequency of the electrons in the applied magnetic field.

The centres of the lines were located by applying a slow linear increase to the magnetic field, and displaying the spectrum on a pen recorder together with proton resonance frequency markers which were arranged to be sufficiently close together to make accurate interpolation possible. Since the electronic g-factor for atomic oxygen has already been measured by Rawson and Beringer (1952), the hyperfine structure interaction constants could be deduced entirely from the separation of the lines, and no attempt was made to measure the absolute fields at which they occur. In spite of the small signal size resulting from the low ^{17}O content, these separations could be measured to within about 0·1 gauss.

The nuclear spin of ^{17}O is 5/2, so that the paramagnetic resonance spectrum of the $^3P_2$ ground state consists of six groups of four closely spaced lines, corresponding to the four transitions for which $\Delta M_J = 1$, $\Delta M_I = 0$ for each of the six values of M_I. At a microwave frequency of about 9580 Mc/s, these groups were separated by roughly 100 gauss intervals. Accurate measurements were confined to those groups for which $M_I = \pm 5/2$ and $\pm 3/2$, the others being insufficiently resolved or obscured by a line due to molecular oxygen. Because the hyperfine structure is much smaller than the Zeeman splitting, the experimental results could be interpreted to a sufficient accuracy using second order perturbation theory. The magnetic dipole and electric quadrupole interaction constants were found to be $a = -218{\cdot}7 \pm 0{\cdot}2$ Mc/s, $b = -10{\cdot}4 \pm 0{\cdot}5$ Mc/s, where a and b are the constants in the Hamiltonian

$$\begin{aligned}\mathscr{H} = {} & g_J \beta H M_J - g_I \beta_N H M_I + ha(\mathbf{I}\,.\,\mathbf{J}) \\ & + hb\left\{\frac{3(\mathbf{I}.\mathbf{J})^2 + \tfrac{3}{2}(\mathbf{I}.\mathbf{J}) - I(I+1)J(J+1)}{2I(2I-1)J(2J-1)}\right\} \\ & - \frac{1}{E_{12}}\left\{\left(\frac{4-M_J{}^2}{12}\right)\left[\beta H - \left(\frac{7+5\kappa}{3-\frac{5}{2}\kappa}\right) ha M_I\right]^2 + h^2 A(M_I, M_J) a^2\right\}.\end{aligned}$$

The last term represents the interaction with the $^3P_1$ state, due to the uncoupling of L and S by the external field and the magnetic dipole hyperfine interaction ; κ is a factor representing the admixture, by electrostatic interaction, of configurations containing unpaired s electrons, and A is a function of M_I and M_J. E_{12} is the separation in zero field of the $^3P_1$ and $^3P_2$ levels.

Before the mean value $\langle r^{-3} \rangle$ for the valence electrons can be estimated from the a constant and the known nuclear moment (Alder and Yu 1951) allowance must be made for the contribution to the magnetic hyperfine structure from configurations containing unpaired s electrons. It would have been possible to evaluate this from the hyperfine structure of the $^3P_1$ state, but these transitions were unfortunately too weak to be observed. However, by comparison with nitrogen, where configuration interaction accounts for the whole of the measured hyperfine structure (Heald and Beringer 1954), the contribution was estimated to be of the order of ± 20 Mc/s. Since this appears in the effect of the interaction between the $^3P_2$ and $^3P_1$ states, the sign of this contribution could be deduced from the experimental results : if the contribution is negative an overall fit with the theory is obtained, which also confirms the order of magnitude of the correction ; this sign would correspond to a positive value of the a constant in ^{14}N. Using, therefore, a value for the magnetic dipole interaction with the $^3P_2$ state of -200 Mc/s, we find

$$\langle r^{-3} \rangle = 3{\cdot}10 \times 10^{25}\ \mathrm{cm}^{-3}.$$

After inclusion of the correction due to polarization of the inner electrons evaluated by Sternheimer (1952), the above value of b is found to correspond (Abragam and Pryce 1951) to a nuclear electric quadrupole moment of $Q = -0{\cdot}0265 \pm 0{\cdot}003$ barn. This error includes the $\pm 5\%$ experimental error in b and a possible $\pm 10\%$ error in $\langle r^{-3} \rangle$. This confirms the value of $-0{\cdot}026 \pm 0{\cdot}009$ found by M. J. Stevenson and C. H. Townes (1957) from the microwave spectrum of HD^{17}O.

We are grateful to Professor B. Bleaney for much helpful advice, and to Mr. L. J. Arundel for valuable assistance with technical problems.

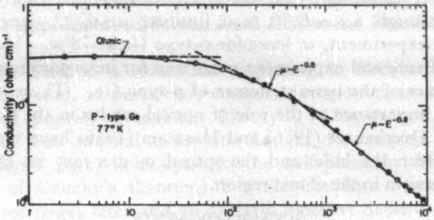
Clarendon Laboratory,
Oxford.
13th June 1957.

R. A. Kamper.
K. R. Lea.
C. D. Lustig.

Abragam, A., and Pryce, M. H. L., 1951, *Proc. Roy. Soc.* A, **205**, 135.
Alder, F., and Yu, F. C., 1951, *Phys. Rev.*, **81**, 1067.
Bowers, K. D., Kamper, R. A., and Knight, R. B. D., 1957, *J. Sci. Instrum.*, **34**, 49
Heald, M. A., and Beringer, R., 1954, *Phys. Rev.*, **96**, 645.
Rawson, E. B., and Beringer, R., 1952, *Phys. Rev.*, **88**, 677.
Sternheimer, R., 1952, *Phys. Rev.*, **86**, 316.
Stevenson, M. J., and Townes, C. H., 1957, *Phys. Rev.*, in the press.

Field Dependence of Mobility in p-Type Germanium

There appears to be considerable uncertainty as to the details and relative importance of the various mechanisms by which carriers in semiconductors can interact with the vibrations of the lattice. In the past, measurements at low fields of the temperature dependence of the mobility of carriers have formed the primary source of data on this interaction. More recently, investigations of the dependence of mobility μ on electric field strength E have been reported; these promise to furnish additional details about the carrier–phonon interaction mechanisms.



The initial work by Ryder and Shockley (1951) and Ryder (1953) on n- and p-type Ge and Si indicated that the field dependence of mobility fell roughly into three ranges: an ohmic range (μ independent of E) at low fields, an intermediate region where $\mu \sim E^{-0.5}$, and finally at high fields a region where the drift velocity (μE) tends to saturate. Recently, much more detailed measurements, with improved sample geometry, by Gunn (1956) on n-type Ge at 300°K showed *no* distinct $E^{-0.5}$ region at intermediate fields, but did show a saturation in drift velocity between 4500 and 9000 v cm^{-1}, followed by a region where $\mu \sim E^{-0.866}$.

The purpose of this note is to describe some results of measurements of the field dependence of mobility in p-type Ge at 78°K and 195°K. Measurements were made on a series of samples with carrier content varying from 3×10^{12} to 4×10^{15} cm^{-3}. The samples were rectangular in shape, and provided with low resistance, non-injecting, soldered end contacts. High fields were applied in the form of pulses, two microseconds long, and with a low repetition rate. The sample resistance was measured in a pulse bridge circuit (Bray 1955) using a difference amplifier and oscilloscope as detector.

The field dependence of conductivity (and hence, mobility) at 78°K of one of the purer samples is illustrated in the figure. There appears to be a gradual variation from an ohmic region to a very distinctive region from 600 to 10^4v cm^{-1} where $\mu \sim E^{-0\cdot 8}$. The slope of $-0\cdot 5$ is shown, but it appears to be tangential to the experimental curve rather than a characteristic feature of it. Saturation in drift velocity (i.e. $\mu \sim E^{-1}$) is *not* observed. The effect of impurity scattering is to increase the field at which deviations from Ohm's law can be observed. For $E \geqslant 10^3$v cm^{-1} all the samples in this study show the characteristic $\mu \sim E^{-0\cdot 8}$ dependence. The conclusion may be drawn that impurity scattering becomes ineffective in the high field region, as expected (Conwell 1953).

At 195°K, the ohmic mobility is much less than at 78°K, and higher fields are necessary to obtain changes in mobility. Again the mobility decreases gradually, showing no extended $\mu \sim E^{-0\cdot 5}$ region. At the highest fields applied (5×10^3 v cm^{-1}) a $\mu \sim E^{-0\cdot 8}$ dependence is not yet apparent.

The hot carrier theory of Shockley (1951) predicts an $E^{-0\cdot 5}$ dependence at intermediate fields for the lattice mobility. This theory, based on a simple energy band structure, takes into account interaction of carriers with the longitudinal acoustical phonons of energy much less than $\boldsymbol{k}T_0$, where T_0 is the lattice temperature. Paranjape (1957) has shown that a limiting $\mu \sim E^{-0\cdot 8}$ relation is attained at sufficiently strong fields when the carriers are hot enough to interact with the acoustical phonons of energy greater than $\boldsymbol{k}T_0$. Thus the complete theoretical picture (based on scattering by acoustical modes *only*) gives a gradual transition from ohmic, through a $\mu \sim E^{-0\cdot 5}$, to a limiting $\mu \sim E^{-0\cdot 8}$ dependence, in fair agreement with experiment, at least for p-type Ge at 78°K. More quantitative comparison of theory and experiment must wait for incorporation into the theory of the peculiarities of the band structure of p-type Ge. There also remains the question of the importance of the role of optical modes in the hot carrier effect. Ehrenreich and Overhauser (1956) and Harrison (1956) have indicated that the interaction between the holes and the optical modes may be the predominant scattering mechanism in the ohmic region.

ACKNOWLEDGMENTS

The authors thank Professor H. Fröhlich for his interest in the work and for communicating Paranjape's results prior to publication. The work was supported by a contract with the U.S. Signal Corps.

Department of Physics,
Purdue University,
Indiana, U.S.A.
28th May 1957.

K. S. Mendelson.
R. Bray.
Communicated by
H. Fröhlich.

Bray, R., 1955, *Phys. Rev.*, **100**, 1047.
Conwell, E. M., 1953, *Phys. Rev.*, **90**, 769.
Ehrenreich, H., and Overhauser, A. W., 1956, *Phys. Rev.*, **104**, 331; **104**, 649.
Gunn, J. B., 1956, *J. Electr.*, **2**, 87.
Harrison, W. A., 1956, *Phys. Rev.*, **104**, 1281.
Paranjape, B. V., 1957, *Proc. Phys. Soc.* B, **70**, 628.
Ryder, E. J., 1953, *Phys. Rev.*, **90**, 766.
Ryder, E. J., and Shockley, W., 1951, *Phys. Rev.*, **81**, 139.
Shockley, W., 1951, *Bell Syst. Tech. J.*, **30**, 990.

REVIEWS OF BOOKS

Physical Mathematics, by C. H. PAGE. Pp. x + 329. (Princeton: van Nostrand; London: Macmillan, 1955.) 42*s*.

It must be emphasized that this is not a book for undergraduates. The opening words of the Preface suggest that the book is addressed to the " average graduate student of physics ", and the author states that his objective is " to help physicists correlate their smatterings of mathematical knowledge in such a way as to improve their feeling for the subject and increase its usefulness ". In spite of the relatively elementary material which makes its appearance here and there, however, the style of the book suggests that the author is really preaching to the converted.

In Chapter One, for instance, entitled " Integration and Differentiation ", the author introduces the Stieltjes integral very reasonably by means of a physical example, but immediately goes on to discuss for a whole page the Riemann and Lebesgue integrals, without defining either of them. In the same place the mathematical conception of *measure* is described, but very inadequately. Indeed, it is difficult to see why this topic was included, for it plays no significant role in the book and will almost certainly confuse the average physicist. Admittedly, in a footnote on page 11 it is pointed out that the *delta function* is not really a function at all, but a measure. None the less, the author's use of the delta function is not appreciably different from that to be found in most physics texts, where this subtlety goes unrecorded.

Later in the book it is somewhat surprising to discover the theory of a complex variable, and, in particular, complex integration, in a section entitled " Application of Fourier Transform Theory ". One cannot escape the conclusion that this was inserted as an afterthought, for on page 211 appears the statement " Hence $F_1(z)$ is analytic in the half-plane by Morera's theorem (the converse of Cauchy's theorem) ", while Cauchy's theorem is not even stated until three pages later and an analytic function defined later still.

It would be a pity, however, if these and other minor irritations should lead to the book being neglected by physicists, for it is in many ways excellent. The average physicist would be well advised to pass lightly over Chapter One and regard the book as essentially a treatise on the differential equations of mathematical physics, with particular emphasis upon Fourier and Laplace transform methods. Eigenvalue problems are very fully considered, and there are useful chapters on function space, integral equations and difference equations. The latter half of the book, dealing mainly with applications of Fourier and Laplace transforms, draws heavily upon electrical circuit theory for its physical examples. This may not appeal to some physicists, but, on the other hand, may prove useful to sophisticated electrical engineers and intending cyberneticists.

Possibly the greatest virtue of the book is its conciseness. A busy physicist might hope to read through it profitably in a finite time, which is more than can be said of some of the compendia of mathematical physics which have appeared recently.

S. RAIMES.

Handbuch der Physik, Vol. 19: *Electrical Conductivity*, I, edited by S. FLÜGGE. Pp. vi+411. (Berlin: Springer, 1956.) DM. 65·60.

In 1933, when the discovery of wave mechanics had made it necessary to supplement the current issue of the *Handbuch der Physik*, it was possible to dispose of the matter in two volumes, entitled *Quantentheorie* and *Aufbau der zusammenhängenden Materie* (in current English usage, ' elementary-particle physics ' and ' solid-state physics '). Some of the articles written then, with their masterly summary of recent achievement, have rightly come to be regarded as classics of scientific literature. Today we face a vastly different situation, and topics which in those days were dismissed with a paragraph or two now form the subjects of massive articles, while the articles themselves have grown into volumes filling a library shelf. And, whereas the year 1933 was a particularly favourable one for taking stock of the new situation in physics, such propitious moments are not likely to recur at present. In the case of solid-state physics, in particular, the foundations are still the same as 25 years ago, but the building activities are on such a big and rapidly increasing scale that reviews of the subject must nowadays be in the nature of reports on work in progress, apt to be quickly overtaken by events; thus another article of the enduring quality of ' Sommerfeld–Bethe ' seems out of the question. With this unavoidable limitation in mind, it can be said that the four articles on solid-state physics in the volume under review (representing, of course, only a few of many in the new *Handbuch*) form a valuable addition to the literature on the subject. The scope of the volume is wider than is indicated by the title, since the longest of the four articles, by J. C. Slater, deals with the general topic of the electronic structure of solids. In addition, A. N. Gerritsen writes on ' Metallic conductivity, experimental part ', H. Jones on ' Theory of electrical and thermal conductivity in metals ', and G. F. J. Garlick on ' Photoconductivity '. (It is, incidentally, another sign of the times that all four articles are written in English.)

Professor Slater's article can be divided into three parts of roughly equal length. The first reviews the periodic potential problem. General aspects of the problem are considered first (this section includes an account of the use of group theory in studying the effect of crystal symmetry), and a second chapter presents detailed methods and results. The second part deals with various other aspects of the one-electron theory of solids, and includes discussions of the basis of the self-consistent field method and of the problem of impurity atoms and alloys. The last part takes up problems such as the crystalline binding energy, the elastic properties, the covalent bond and the magnetic properties of solids, in which the interaction between electrons is of decisive importance. The most difficult task here is to take proper account of correlation effects, and various alternative ways of doing this are described. Professor Slater has succeeded in treating an impressive variety of topics and his success is all the more remarkable in that his treatment, while concise, is never superficial. His method is to confine himself strictly to the essential ideas involved in any argument and to avoid formal mathematics almost entirely (the article contains an average of 0·7 equations per page). Indeed Professor Slater's style is so admirably clear and persuasive that one hopes it will cause no one to underestimate the difficulties of an inherently intricate subject.

Professor Gerritsen's article gives a competent account of the main experimental facts on the electrical conductivity of metals at normal and at high temperatures, and brings up to date the older accounts by Grüneisen and Meissner. After a chapter devoted to definitions and a survey of methods of measurement, he discusses in successive chapters the observational data in relation to the effect of electron configuration and crystal structure, temperature, magnetic fields, pressure, crystal defects and alloying. His review brings out clearly the extremely sensitive dependence of the electrical conductivity on the physical state of the metal, and it is evident that the subject, which bristles with anomalies, will present many problems to theory for a long time to come.

The framework of the conduction theory is laid down in the first two chapters of Professor Jones's article. In subsequent chapters he examines in detail the consequences of the theory for particular models, discussing in turn the wave mechanics of the scattering of electrons, the conductivity of impure metals and alloys, the thermal conduction effects, conduction at low temperatures, magnetic effects and finally the anomalous skin effect. It is good to find that the treatment includes careful accounts of the role played by Umklapp processes in the conduction process, and of the use of the variation principle for obtaining solutions of the transport equation. Professor Jones, in dealing with a much more restricted subject than Professor Slater, has wisely adopted an entirely different approach: he develops the mathematical argument in full detail (the equation density is about 8 times as high as in Slater's article), but takes great care to state as precisely as possible the assumptions adopted at each stage of the calculation. The question arises (and forms a well-known bone of contention) as to just how seriously one should take the results of the theory, bearing in mind that in most cases very drastic simplifying assumptions have to be introduced, especially in the final stages of the calculations. I feel that, on occasion, Professor Jones is a little over-optimistic in his estimate of the ability of the theory to account for the anomalies in the experimental facts. I was also sorry not to find any discussion of the possible use of phenomena such as the anomalous skin effect in studying the form of the Fermi surface in real metal crystals; but these remarks are perhaps no more than expressions of personal prejudice.

Professor Garlick's article deals with a subject which, except for the classic experiments of Gudden and Pohl, had remained largely unexplored up to the time of the last war, and his main concern is to review the enormous developments of the past few years. After a brief historical introduction he describes the theoretical concepts—energy bands, trapping states, barrier layers and so on—which are essential for the interpretation of the experimental data. He then reviews the data for individual substances, dealing first with monatomic solids and then, in the main chapter, with the large class of binary compounds which show photoconductivity. The article concludes with brief notes on photoconductivity in organic systems, on the photodielectric effect in solids and on some recent experimental techniques in photoconduction measurements. Professor Garlick's account provides a useful guide to an extensive amount of work, much of which has been published in relatively obscure or inaccessible journals.

All in all, the purchase of Volume XIX of the new *Handbuch* is strongly recommended to all scientific libraries, whether or not they feel rich enough to acquire the whole series, and many research workers will also wish to possess their own copy.

E. H. SONDHEIMER.

Robert Hooke, by MARGARET 'ESPINASSE. Pp. vii + 192. (London: Heinemann, 1956.) 21*s.*

All who have occasion to dip into the history of the formative century of modern science in almost any field—biology, physiology, chemistry, astronomy, physics—are bound to come across the work of Hooke which, as often as not, includes the decisive first step of systematic observation and experiment. In the history of scientific institutions he more than any other, as Curator of Experiments for the new Royal Society, provided the model for the scientific research of the future. Because he was so many sided it is difficult to see his life as a whole, though after the publication of his diaries we know more about his life than any other seventeenth century scientist. We are therefore particularly grateful to Mrs. 'Espinasse for her all-round study of Robert Hooke which presents him in the social environment of his time and helps to show the reasons both for his successes and his failures.

The failures were not in science, for he succeeded so well that in many fields his work was not bettered for a hundred years, but in getting due recognition for his work. The neglect of Hooke's achievements Mrs. 'Espinasse puts down mostly to the determined spite of Newton, under whose influence science in Britain took on for a century a highly mathematical, philosophical turn quite foreign to the lively practical interests of Hooke. But all through his life Hooke suffered, although for the most part he took it very well, from having to be second to other men, wealthier, more pushful and only scarcely more talented than himself. He started as an assistant to Boyle, and science owes the air pump and ' Boyle's law ' to him. In the Royal Society he suffered for years the animosity of the Secretary, Oldenburg. He saw his invention of the balance spring clock taken by Huyghens after some very dubious procedures on which Mrs. 'Espinasse throws much light. His greatest theoretical achievement, the enunciation of the inverse square law of gravitation, which he may well have done before Newton, only brought him trouble. Finally, in his architectural work he followed a most willingly acknowledged master, Wren. To be second to such giants is itself a title to greatness but Hooke certainly deserves more than that. The fame that is only now coming to him he missed in his time because he did not seek it; he was more interested in understanding and mastering nature and art than making a great name and place for himself. What he actually did for science is not to be found in any of his works taken separately, but rather in the general impulse he gave to experimental science, building up a tradition which, whether it acknowledged its debt to Hooke or not, has continued and flourished to this day. It still remains with its insistence on the primacy of experiment and its distrust of a prior theory specially characteristic of British science.

Mrs. 'Espinasse's book, admirably illustrated from the *Micrographia* and his other works, would furnish to those who know little of Hooke, except Hooke's Law or Hooke's universal joint, an admirable picture of the man in the round; to those to whom he is already a favourite she presents a new aspect of his personality and has many penetrating things to say. Not everyone will agree with her judgment of Newton but we cannot doubt that he failed to give Hooke his due.

J. D. BERNAL.

Laboratory Administration, by E. S. HISCOCKS. Pp. xvi + 392. (London: Macmillan, 1956.) 36*s*.

In this book the everyday problems which confront anyone responsible for the administration of a large modern laboratory are discussed in an easily readable manner.

The author describes and compares the methods used in a number of government and industrial laboratories in this country and abroad for dealing with the recruitment, assessment, remuneration and training of staff, with the provision of buildings and services, the flow of information, and with the programming, control, and evaluation of laboratory work.

Prominence is naturally given to the practices followed at the N. P. L. where Mr. Hiscocks is Secretary, but the book also contains much information culled from other sources. Many comparative statistics and examples of rating forms and organization charts are set out in 22 Appendices and together with a bibliography provide a useful source of reference.

Mr. Hiscocks has provided a broad survey of current practice in a field of management which is still developing. He does not claim that any of the administrative practices which he describes are the best or final answer to the problems discussed, and the book contains little that is really new. T. D. S.

Matrix Calculus, by E. BODEWIG. Pp. xii + 344. (Amsterdam: North Holland, 1956.) 53*s*.

The subject matter of this book is linear algebra—the solution of simultaneous linear equations, the inversion of matrices, and the determination of latent roots and vectors. It would be difficult to think of any branch of mathematics which has a wider range of application: all branches of engineering and applied physics and applied statistics depend heavily on it. It is a branch of mathematics which exists at several levels: at the foundation is the basic theory, as taught to, say, second year undergraduates, and on this are erected layers concerned with numerical problems arising from practical applications of the basic theory, in successively increasing degrees of sophistication. These are associated with 'failing cases'—the difficulties which arise in the attempted inversion of a nearly singular matrix, for example, or the evaluation of the vectors associated with nearly coincident latent roots.

It will be readily appreciated that this is a subject whose study has been greatly accelerated by the availability of electronic computers. Large stocks of results and methods are coming forward from all sides, and what one needs now is a critical comparative survey by an author whose all-seeing eye enables him to unify the mass of material available, without ever allowing his work to degenerate into a mere catalogue.

This is admittedly asking a lot. There are parts of Bodewig's book which are pure catalogue, but it must be admitted that there are also parts where he has succeeded in giving a critical unifying survey, notably in his discussion of the various iterative methods of solving linear equations. The quality of the writing fluctuates between these two levels throughout the book. Even the catalogue level, however, is of value in that the results listed are mainly recent and often difficult to access elsewhere, many having previously appeared only in the original papers. Of the 111 references cited in the bibliography, about two thirds are of post-war date. There is also a useful index.

The book is divided into four sections. Part I, Matrix Calculus, gives the account of basic matrix algebra; Part II deals with the solution of linear simultaneous algebraic equation, including a brief discussion of ill-conditioned systems, Part III with the inversion of matrices, and Part IV with eigenvalue problems, both direct and iterative methods being discussed throughout.

One slightly unexpected item is the detailed discussion of ' geodetic matrices ', that is to say, matrices of the type which arise in the normal equations of least squares obtained from the reduction of triangulated survey data.

An aspect of the book which invites comment is the Dutchness of the English in which it is written. It is probably fair to say that at no point is it impossible to penetrate to the author's meaning, but his usages are often confusing, being nearly enough right to have *a* meaning, but not quite the meaning intended. It has evidently been written direct in English by the author (a praiseworthy achievement) and never subsequently vetted.

One other idiosyncratic aspect of the book is the author's refreshing and unashamed enthusiasm for his notation. This depends on a consistent use of matrix concepts, with free use of partitioning, and with the basic building bricks of a matrix taken to be its vectors (rows or columns) and not its scalar elements. This is in the reviewer's opinion the correct way to teach and use matrices, though it is perhaps not as novel as the author claims.

To conclude, the book furnishes an adequate account of the basic theory and a reasonably complete summary of recent developments over the whole field of practical linear algebra. Workers having problems of numerical evaluation in this field will find it an extremely useful handbook. E. H. LLOYD.

Fifth Symposium (International) on Combustion. Combustion in Engines and Combustion Kinetics. Published for the COMBUSTION INSTITUTE. Pp. xxvi + 802. (New York: Reinhold; London: Chapman and Hall, 1955.) 120*s*.

This volume is a collection of more than one hundred papers presented at the Fifth Symposium on Combustion which was held at the University of Pittsburgh, from August 30 to September 3, 1954. The field of combustion is now so large that it is advisable at a meeting of this kind to concentrate on some particular aspect of it and on this occasion the organizers focused attention on combustion in engines and combustion kinetics. At the beginning of the book there are eleven survey papers, six of which deal with combustion problems in internal combustion engines, diesels, liquid fuel rocket engines, solid propellent motors, turbo jets and ramjets respectively. Each is written by an author well known for his experience with a particular type of engine. The remainder of the papers in this group are more concerned with the fundamentals of combustion phenomena and two of these are of special interest because the authors challenge the validity of applying conventional kinetic theory to flame reactions. The first, on the kinetics of elementary gas phase reactions at high temperature by Kurt Schuler has as its underlying theme the importance of the large amount of energy which is added to and contained in the reaction system, and the way in which the presence of rather large concentrations of high energy species modifies the reaction mechanism and kinetics. The existence of excited species with high energy content is now beyond doubt but the importance of energy chains as compared with radical chains cannot be decided until more facts are known and the author indicates ways in which basic research could help. The second paper, by Professor A. R. Ubbelohde, is a survey of collision processes in the combustion of hydrocarbons in which the physico-chemical concepts of the kinetic theory of the velocity of gas reactions when applied to polyatomic molecules are discussed. Whether polyatomic molecules or large skeletons derived from them exist for an appreciable time in high temperature reactions is not known, and this paper too shows the need for some basic experimentation. In his introductory remarks at the Symposium, Professor Hottel emphasized the importance of the survey papers and expressed a hope that those engaged in fundamental studies would pay attention to them. How well his remarks will be heeded in the future it is impossible to say but it is perhaps profitable to bear them in mind when one reads the individual research papers which make up the bulk of the volume. These papers are grouped as follows; the figures in parenthesis are the number of papers in each group: Combustion of fuel droplets (6), Propellent burning (9), Combustion of solids (6), Diffusion flames and carbon formation (5), Combustion in engines (8), Special techniques (5), Kinetics of combustion reactions (45), Flame spectra and dissociation energies (6).

In the large group on kinetics there are a number of papers dealing with the oxidation of the simpler hydrocarbons and aldehydes. The investigations are usually made at temperatures which are convenient in the laboratory and lead to manageable rates of change. The results are only of very limited application in combustion. Studies of this kind are not new and are not always very rewarding. The mechanisms put forward offer a satisfactory explanation of the results obtained but necessarily contain steps involving free radicals about which there is usually very little definite information. Worth mentioning in this connection is a paper on the thermochemistry of the alkoxy radicals, RO, by Peter Gray in which the heats of formation of the lower alkoxy radicals are deduced and used to derive some of the bond dissociation energies in compounds containing the alkoxy groups. At the present time one would like to see more papers of this kind. A welcome feature of this section are a number of papers dealing with fuels other than hydrocarbons and oxidizers other than air. Work on propellents and explosives has led to a broadening of the combustion field and some interesting comparisons may be drawn between the different systems.

In some of the smaller groups there are papers dealing with the burning of liquids and solids and with studies of combustion in engines. In many of these the chemical reaction rates are either ignored or shown to be of little importance compared with the fluid dynamics of the system and considerations of heat and mass flow. There is almost as much to interest the physicist as the chemist in these groups and in a good group of papers dealing with flame spectra. At any symposium one is keen to know about new methods of investigation and at this meeting there were five papers presented on special techniques. One of these papers by H. S. Glick, W. Squire and H. Hertzberg describes a modified shock tube method for the study of high temperature gas phase reactions in which the conventional shock tube set-up has been altered so as to obtain a single high temperature pulse of known shape and duration.

To sum up, this book contains contributions from the majority of laboratories working on combustion and is therefore of great interest to any one who labours in this field. For others it provides survey papers, from which they can quickly learn of some of the problems which have arisen in the development of various types of engines and of the limitations of knowledge, particularly in the kinetics of reactions at really high temperature (1000–3000°c). It is perhaps a pity that very little of the discussion at the Symposium is included but this might have led to a volume of prohibitive cost. W. G. PARKER.

Théorie et technique de la radiocristallographie, By A. GUINIER. Pp. xviii + 736 + XVI. (Paris: Dunod, 1956.) 9500 *fr*.

This book is, in origin, a second edition of *Radiocristallographie*, which appeared in 1945, and was written without much knowledge of non-French work later than 1940. An English translation, with some revision, was issued in 1952 (reviewed in the *Proceedings of the Physical Society* B, **65,** 916).

The new book is about twice the size of the old, and the great development of the subject since 1945 has made it necessary to rewrite all but a few paragraphs. The confusion between ‘ lattice ’ and ‘ structure ’, complained of in the review cited above, has been rectified, and F_{hkl} and A_{hkl} are clearly distinguished, though the last does not seem to be mentioned by name. The scope is limited in the same fashion; the determination of crystal structures is barely mentioned (nine pages out of over 700), but apparatus, techniques, and the theory of scattering are treated in considerable detail. “ We have had in view the needs of other users of x-rays in the laboratories, not only [laboratories] of crystallography, but of physics of solids, of chemistry, of metallography, etc. . . . One is working, then, with crystals whose structure is known or deliberately left undetermined. The problems presenting themselves are, for example, phase identification, quantitative analysis of a solid, mode of crystallization, imperfections of structure, etc. . . .”

Of the 700-odd pages in the book, 72 are devoted to general properties of x-rays, including sources and detectors, 82 to elementary crystallography and diffraction theory, 148 to experimental methods, 98 to applications, and 261 to diffraction by imperfect crystals and amorphous materials. The chief technical advance recorded since 1945 is the wide application of counter methods, and the chief theoretical advance is in the interpretation of diffraction by imperfect crystals. The latter development has led to the abandonment of the elementary mathematical level of the earlier work. Much use is made of Fourier transforms, particularly in the last part. An Appendix (25 pages) is devoted to them. Each of the 14 chapters is followed by a bibliography, and there is a further bibliography of general works. The book ends with about 30 pages of tables, and the indexes are more extensive than is usual in French books.

Comparison with an English book of similar scope is interesting. Guinier devotes some 50 pages to single-crystal techniques (deliberately excluded from the English book), and one page to high- and low-temperature cameras (36 pages in the English book). As would be expected, his treatment of low-angle scattering is very thorough. Such differences in emphasis make Guinier's book essential to serious workers in the field of diffraction by polycrystalline materials.

The treatment is always lucid, the figures clear, the accuracy high. The book can be criticized seriously only on the ground of price (nearly £10), and perhaps weight (over 3 lb, which means that it cannot be read comfortably in an armchair, but only at a desk). A minor idiosyncrasy is the figure numbering, consecutively throughout each chapter in the form Fig. V-1, Fig. V-2 etc. in the figure legends, but 5.1, 5.2 etc. in the text. There are occasional misprints, such as a displaced square bracket on p. 112 and ' Stoks ' for Stokes. None noticed by the reviewer is likely to cause any difficulty. A. J. C. WILSON.

Surveys in Mechanics, edited by G. K. BATCHELOR and R. M. DAVIES. Pp. vii + 475. (Cambridge: University Press, 1956.) 50*s*.

This recent addition to the Cambridge Monographs on Mechanics and Applied Mathematics was published to commemorate the seventieth birthday of Sir Geoffrey Taylor. It takes the form of a biographical note by R. V. Southwell followed by ten separate articles, each of which is a survey of the present position in some field of mechanics in which G. I. Taylor has been active at some time during the past fifty years. Every article is written by a recognized authority on the subject in question and is intended to be suitable for the non-specialist.

By its very nature the framework of the book is built around the immense activity and scientific output of the man to whom it is dedicated. Beyond this common feature no attempt has wisely been made to link the articles together. The field which is covered is at the same time both wide and restricted. It is extensive in the sense that the articles range from such widely diverse subjects as ' Turbulent Diffusion ' by G. K. Batchelor and A. A. Townsend to ' Dislocations in Crystalline Solids ' by N. F. Mott. On the other hand each selected topic is treated in detail up to the limits of present knowledge.

The emphasis on the whole is on the application of mathematical analysis to the understanding of experimental systems as, for example, in the commendable article on the ' Mechanics of Sailing Ships and Yachts ' by K. S. M. Davidson or in that by F. Ursell on ' Wave Generation by Wind '.

In addition to the subjects already mentioned the ' Mechanics of Quasi-static Plastic Deformation in Metals ' are reviewed by R. Hill and the ' Mechanics of Drops and Bubbles ' by W. R. Lane and H. L. Green. H. B. Squire contributes a mathematical treatment of ' Rotating Fluids ' or, more specifically, on the motion of a body in a fluid which has a uniform angular velocity far away from the body. Problems associated with wave propagation, including waves of infinitesimal amplitude and shock waves, are dealt with by two authors: ' Stress Waves in Solids ' by R. M. Davies and ' Viscosity Effects in Sound Waves of Finite Amplitude ' by M. J. Lighthill. Both these articles are lengthy and provide excellent summaries of the respective fields.

A notable feature of the series of articles is the excellent composition and clarity of expression which is maintained throughout. As would be expected in review papers, adequate references are provided for each subject. In some cases, particularly, a list of symbols and of definitions would have helped the non-specialist reader.

Undoubtedly, this is a fitting tribute to a great scientist. The book can be thoroughly recommended to workers in any field of mechanics as presenting a lucid account of the state of knowledge in related subjects. Owing to the very diversity of the subjects treated combined with the detailed treatment of each selected topic, it is probable that many general readers, not specifically working in the field of mechanics, will prefer to obtain the book from a reference library rather than to possess an individual copy. J. LAMB.

Handbuch der Physik, Vol. 1, *Mathematical Methods I*, edited by S. Flügge. Pp. vii + 364. (Berlin: Springer, 1956.) DM. 72.

The first two volumes of the new edition of the *Handbuch der Physik* are on Mathematical Methods. The first volume contains five articles, all in German. Three are by Lense on " Basic concepts of classical analysis, ordinary differential equations, theory of functions ", " Partial differential equations " and " Elliptic functions and integrals "; one is by Meixner on " Special functions of mathematical physics ", and one is by Schlögl on " Boundary value problems ".

The introductory article by Lense gives a clear and compact account of many of the basic concepts and facts of classical calculus. Definitions are given, theorems are stated in a logical order, and a vast amount of material is compressed into a relatively small space. Most of the text is easy to read and will be useful for reference to the physicist who wants only to refresh his memory or to inform himself mainly about facts. Real functions with one and several variables, differential and integral calculus, and infinite series are dealt with. One section reviews the theory of functions of a complex variable, and a very clear section gives an account of the basic facts of ordinary differential equations. An appendix deals with the Lebesgue integral.

Lense's article on partial differential equations (which is rather different from the same author's contribution in the 1928 edition) deals mainly with equations of the first order. Second order equations reappear in the other articles, but the subject is not adequately covered by the volume. The concepts related to the solutions of first order partial differential equations, their geometric interpretation and the theory of contact transformations are first explained for the more readily visualizable case of equations with two independent variables. The considerations are then extended to several variables and the connection with the Hamilton–Jacobi theory is established. The exposition is too concise to serve as a first introduction into the subject, but is suitable for reference.

Lense's third article gives a clear account of many of the features of elliptic functions and integrals. The exposition is descriptive, and in the nature of a summary, but this article can be read also as a first introduction. It presents in a nutshell the concepts and relationships of a special topic and gives useful orientation in a rather extended field.

Meixner reviews some of the properties of special functions, the hypergeometric function with its special and limiting cases (Bessel functions, spherical harmonics, Whittaker functions), the Matthieu functions and the spheroidal functions. The emphasis is on the general methods by which these functions can be handled. Information concerning a particular function therefore appears scattered throughout the paper. (When these special functions are actually to be used, one would rather consult the existing monographs). Definitions and properties, the related second-order linear differential equations and functional equations are treated; the account of the latter includes the factorization method and some of the main results of Truesdell's unified approach. Some references to recent contributions of physical interest are given. The separation of the three-dimensional wave equation in different curvilinear coordinates is presented in a systematic way, the differential equations of the special functions appearing through the separation.

Schlögl's article on boundary value problems deals with systems of orthogonal functions, integral equations, variational calculus and the boundary value problems of the differential equations of physics. A great many concepts are reviewed and there is more stress on the explanation of the concepts than on technical details. This would be an advantage, but for an introduction into the subject the article remains too summary. Concepts related to function spaces and expansions, Fourier series and integrals, and theorems and formal methods of solution for integral equations are reviewed. Boundary conditions for partial differential equations of the second order, characteristics, Green functions and eigenvalue problems are dealt with. Propagators of initial value problems are given for some simple cases.

The classical topics of mathematics which form the content of this volume are treated in many books, and the volume cannot be expected to compete in all its details with the existing expositions. It has its merits, however, and gives a handy collection of information which will be welcomed by the physicist.

J. G. VALATIN.

Elements of X-Ray Diffraction, by B. D. CULLITY. Pp. xiv + 514. (Reading, Mass.: Addison-Wesley; London: Academic Books, 1956.) 80*s*.

The author is a metallurgist, but states that his book should be useful also for chemists, physicists, ceramists and mineralogists. It is divided into three parts: fundamentals (137 pages), experimental methods (Laue and powder photographs, diffractometer, 78 pages), applications (240 pages); and ends with some mathematical appendices, tables of functions and physical constants. There are numerous problems, some with answers. A short chapter on suggestions for further study mentions the main textbooks, reference books and periodicals, but there are no references in the text.

The book is avowedly an elementary introduction, and as such may have its uses. The author writes with an obvious enthusiasm for his subject that atones for some of the shortcomings of the treatment, which are many and minor. For example, the angle of view of perspective drawings is such that two body diagonals of a cube coincide. The fourfold inversion axis is explained with reference to a cube, in which it is masked by a more obvious fourfold axis. In the AuBe structure (p. 49) not only the Au but also the Be atoms are displaced from the face-centred-cubic positions. The five unit cells of fig. 2–21 (*a*) contain ten atoms according to the rules given on p. 33, not 29 as stated in the text. " No known element crystallizes with a simple cubic lattice containing one atom per unit cell " (p. 50), but on p. 483 this structure is attributed to polonium at ordinary temperatures. A beginner can straighten out a few such confusions, but their cumulative effect retards his progress. A. J. C. WILSON.

CONTENTS OF SECTION A

PAGE

Dr. E. W. Laing and Dr. R. G. Moorhouse. Photoproduction of Pions at Complex Nuclei 629

Dr. W. D. Allen and Dr. A. T. G. Ferguson. The Measurement of Fast Neutron Flux over the Neutron Energy Range 0·03 Mev to 3·0 Mev 639

Mr. T. E .Peacock. Electronic Absorption Spectra of Heterocyclic Systems II: Applications to some Two-ring Nitrogen Heterocyclics 654

Dr. P. C. Price. The Radiative Capture of Alpha Particles in ^{15}N 661

Dr. G. E. Brown and Dr. C. T. De Dominicis. Elastic Scattering of Low Energy Nucleons by Complex Nuclei 668

Dr. G. E. Brown. Resonances in a Complex Well 681

Dr. G. E. Brown and Dr. C. T. De Dominicis. Direct Interaction and Nuclear Dispersion Theory 686

Dr. T. V. Kanellopoulos and Dr. G. E. Brown. Polarization Experiments in Proton–Proton Scattering 690

Dr. G. E. Brown and Dr. T. V. Kanellopoulos. Analysis in Coincidence in Proton–Proton Scattering 703

Corrigenda:

Dr. J. Bowcock. Compound Nuclear Theory and the Optical Model . 708

Reviews of Books 709

Contents of Section B 712

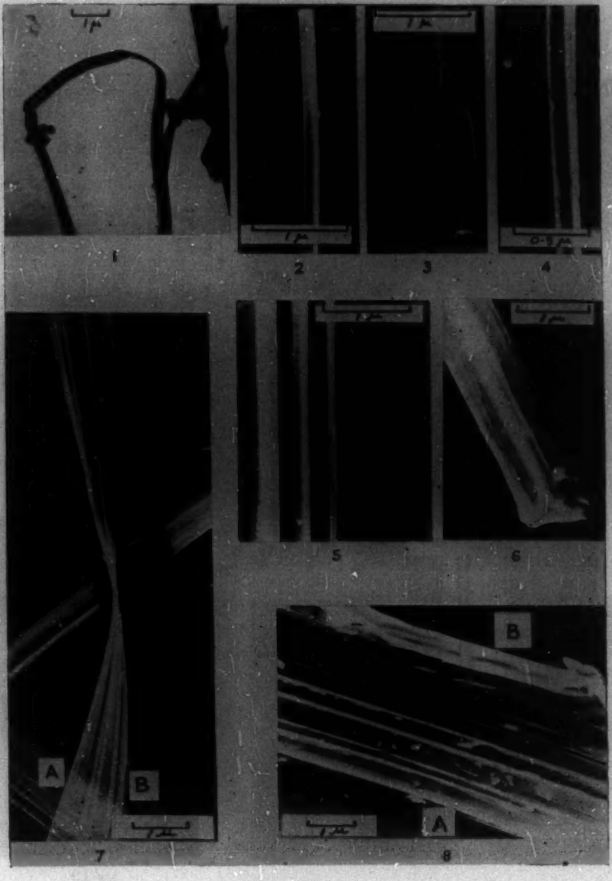


Plate I

Figure 1. Direct electron micrograph of a tin whisker in the form of a ribbon. ×6000.

Figure 2. Shadowed carbon replica of a smooth, rod-shaped whisker. ×17 000.

Figure 3. Shadowed carbon replica of rod-shaped whiskers branching. ×17 000.

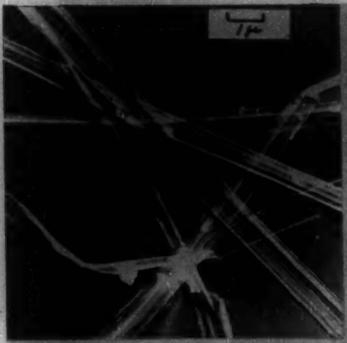
Figure 4. Shadowed carbon replica of a whisker with a star-shaped cross section. ×28 000.

Figure 5. Shadowed carbon replica of a coarsely-ridged whisker. ×14 000.

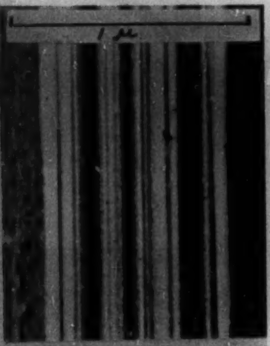
Figure 6. Shadowed carbon replica of a ribbon with a nearly triangular cross section. ×12 500.

Figure 7. Shadowed carbon replica of a ribbon-shaped whisker. ×9000.

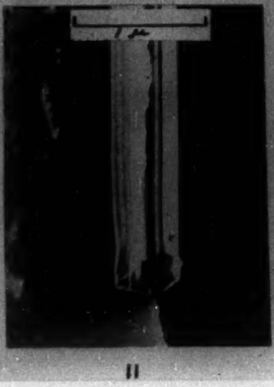
Figure 8. Shadowed carbon replica of a large ribbon-shaped whisker. ×8000.



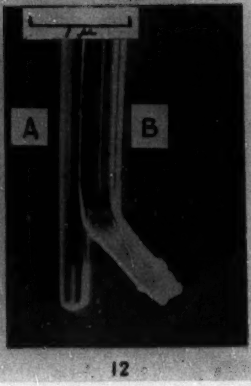
9



10



11



12

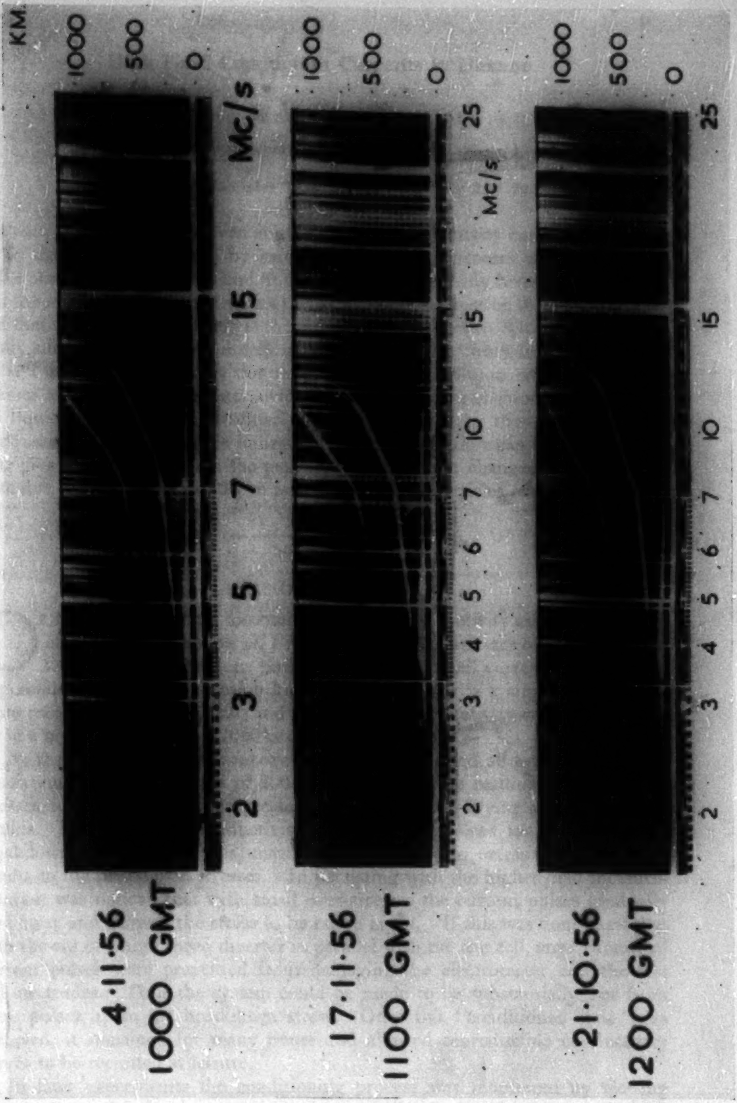
Plate II

Figure 9. Shadowed carbon replica of a number of ribbon-shaped whiskers. ×6500.

Figure 10. Shadowed carbon replica of a whisker with many fine ridges. ×44 000.

Figure 11. Shadowed carbon replica of a whisker which resembles a stick of celery. ×24 000.

Figure 12. Shadowed replica of a whisker which has split at the end. ×18 000.


KM.
1000
500
0
Mc/s
25
15
10
7
6
5
4
3
2
4·11·56
1000 GMT
17·11·56
1100 GMT
21·10·56
1200 GMT

High Field Conduction Currents in Hexane

By H. HOUSE

Electrical Engineering Department, Queen Mary College, London

MS. received 20*th November* 1956, *and in final form* 21*st June* 1957

Abstract. The author has been able to measure conductivity currents in hexane up to the breakdown stress by using a conditioning process which enables a stable state of the electrodes and liquid to be kept for many hours. It is shown that the results are consistent with a theory of electron emission from the cathode, and that the dominant process is the Schottky one, together with the mathematically similar barrier peak tunnelling component. The work function is found to be 1·0 to 1·2 ev, although this may contain a bulk liquid component. For stresses of 1 MV cm^{-1} and greater, evidence is given of an ionization process within the liquid. Due to the difficulties of analysis, caused by the use of spherical electrodes, no estimate of the ionization increment factor α can be given. The tests give no information on the possibility of transient changes in the cathode state due to its inability rapidly to neutralize ionized particles reaching it.

§ 1. INTRODUCTION

OF the many attempts to obtain high field conductivity curves for simple organic liquids, almost all have been limited to stresses of a few hundred kv cm^{-1} by the spurious pulses which make small current measurement impossible. The only exception known to the author is a single result given quite recently by Green (1956) and this shows a conductivity curve for hexane up to a breakdown stress of 1060 kv cm^{-1}.

At the outset of the present work, the liquids tested allowed conductivity measurements up to stresses of 200 to 400 kv cm^{-1}, the readings at the higher stresses being attainable when greater care was taken in drying and filtering the liquids. The high field conductivity values were required for a study of the breakdown process, as it was unsatisfactory to have to extrapolate low stress results to the breakdown stresses. In persisting with the higher field measurements it was noticed that with small over-stresses the current pulses gradually died away and allowed the stress to be raised again. If this was done carefully, with the aid of a high speed diverter in parallel with the test cell, any ‘ runaway ’ current pulses were prevented from damaging the electrometer and the test cell electrodes. Thus the system could be made to be substantially free from these pulses up to the breakdown stress. Once this ‘ conditioned state ’ was achieved, it remained for many hours and allowed reproducible conductivity curves to be recorded at leisure.

In later experiments the conditioning process was monitored by viewing the greatly amplified pulses on a cathode-ray oscillograph. As this gave a much better control of the process, the operation of the diverter was avoided.

§ 2. EMISSION FROM THE ELECTRODES INTO THE LIQUID

2.1. *Electron Emission Processes*

At the surface barrier of a metal, electrons are emitted by two main processes: by passing over the barrier or by tunnelling through it. Figure 1 (*a*) shows a general potential barrier which electrons have to overcome before they can leave the metal. Figure 1 (*b*) shows, for normal temperatures in an ideal metal, the 'energy density' of the electrons at the various values or levels of energy. The number of electrons of a given energy which will escape from the metal will depend upon the electron density corresponding to that energy and the ease with which the barrier may be passed at this energy. Since it is not practicable to derive a single mathematical expression for the total number of electrons, of all values of energy, which would pass from the metal, it is usual to consider separately the electron flow from each of a number of energy ranges. The relative contribution which each of these ranges will make to the total emission current is dependent upon the temperature and the shape of the energy barrier. The barrier shape will in its turn be determined by the applied field F_a, the work function $e\phi$, the type of image law, surface dipole layers due to absorbed molecules and surface space charge phenomenon in the materials close to the metal surface.

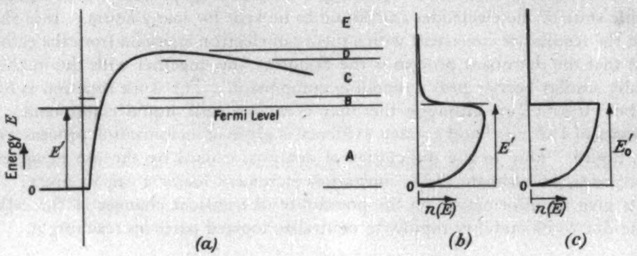


Figure 1. The energy barrier of the surface.

In the following paragraphs an assessment of the magnitude of the electron emission current, for a normal image type barrier with a work function value of 1·2 ev (this value being of the order suggested by the experimental work) will be made for each energy range and for several values of the applied field F_a.

Energy Ranges A–B

The electrons in the energy range A are those with energies less than the Fermi level. An expression for their escape from a metal was first given by Fowler and Nordheim (Fowler and Nordheim 1928, Nordheim 1928) but the equation below is in the form given by Sommerfeld and Bethe (1934):

$$j_A(F_a) = 1{\cdot}55 \times 10^{-6}(F_a^2/e\phi) \exp\{-6{\cdot}86 \times 10^7 (e\phi)^{3/2} v(y)/F_a\} \qquad \ldots\ldots (1)$$

where $j_A(F_a)$ is the Fowler–Nordheim emission current density in A cm^{-2}, $e\phi$ the work function in ev (measured when $F_a = 0$), F_a the applied field in v cm^{-1}, $v(y)$ a function of y and $y = 3{\cdot}62 \times 10^{-4} F_a^{1/2}/e\phi$.

A table of values of $v(y)$ has been given in an article by Burgess, Kroemer and Houston (1953).

Table 1 shows values of $j_A(F_a)$ for several values of F_a.

Table 1. Fowler–Nordheim Emission Currents

F_a (MV cm^{-1})	1	1·2	1·5	2	3
$j_A(F_a)$ (A cm^{-2})	$1{\cdot}2\times10^{-28}$	$3{\cdot}9\times10^{-22}$	$2{\cdot}2\times10^{-15}$	$5{\cdot}8\times10^{-9}$	$2{\cdot}9\times10^{-2}$

The emission current density into a liquid will be smaller than the vacuum value given in table 1, since the changed image law will result in a wider barrier.

The Fowler–Nordheim (F–N) equation is derived for a temperature of 0°K, so that the electrons have energies $E < E_0'$ and are distributed in the manner indicated in figure 1 (*c*).

At higher temperatures the 'Fermi tail' of the electron energy distribution extends into the higher energy regions; the density decreasing in an exponential manner as the energy E increases. Due to the changed distribution of the electrons at the higher temperatures, the Fermi level rises a little and in consequence the electron emission for energies less than the Fermi level E' is increased (Guth and Mullin 1942). For $F_a = 1$ MV cm^{-1} and $T = 300$°K the emission is almost double the F–N value.

No expressions, that are valid for $F_a = 1$MV cm^{-1} and $T = 300$°K, are available for the electron emission in the range B. Its value would be expected to be of the order of the F–N value.

Energy ranges E *and* D

The electron emission from the energy range E is given by the Schottky expression for field aided thermionically emitted electrons.

$$j_E(F_a) = AT^2\bar{D}\exp\{-(E_{zm} - E')/\boldsymbol{k}T\}$$

where $A = 120{\cdot}4$ A cm^{-2}, $E_{zm} - E' = \boldsymbol{e}\phi - \boldsymbol{e}(\boldsymbol{e}F_a/\epsilon_r)^{1/2}$, T = temperature of surface in °K, $\bar{D}$ = average barrier transmission coefficient for the energy range E in figure 1 (*a*).

Table 2 gives the values of $j_E(F_a)$ for $T = 300$°K and $\bar{D} = 1$.

Table 2. Schottky Emission Currents

F_a (MV cm^{-1})	1	1·2	1·5	2	3
$j_E(F_a)$ (A cm^{-2})	$1{\cdot}8\times10^{-10}$	$5{\cdot}1\times10^{-9}$	$2{\cdot}3\times10^{-8}$	2×10^{-7}	$7{\cdot}7\times10^{-6}$

These values are calculated for the liquid-modified image curve, which causes a smaller Schottky reduction than the vacuum image law. The values of $\bar{D}$ are much smaller at normal temperatures (300°K) than at the temperatures used for thermionic valves. Typical values of $\bar{D}$ corresponding to Schottky emission into a vacuum are given in Appendix B.

Guth and Mullin (1942) have derived an expression for the emission of electrons in the energy range D. The current $j_D(F_a)$ of electrons which tunnel through the relatively narrow peak of the barrier when the applied stress is 1 MV cm^{-1} is about equal in value to the Schottky emission current.

It is thus seen that, for the ideal conditions and a work function of 1·2 ev, almost all the electrons emitted at stresses of less than 2 MV cm^{-1} are from the Schottky emission region E and region D. Guth and Mullin have shown that the emission from the region D has the same law as the Schottky emission multiplied by a slowly varying function of the applied field F_a. Thus for the present work it is convenient to assume the Schottky emission law and use the Schottky plot to detect other effects by the changes they may cause to the expected straight line emission curve. The slope of the experimental Schottky plot will be modified by the barrier peak tunnelling component and also by the variations in the value of $\bar{D}$, calculated in Appendix B.

2.2. *The Electrode Surface*

The values of the emitted current determined in § 2.1 assume smooth electrode surfaces of uniform work function.

Surface irregularities cause local increases in the electric field at the extremities of the sharper points, but these stresses decrease extremely rapidly on leaving the surface. The relative change in the emission will be different for the Schottky and Fowler–Nordheim processes and will be dependent upon the mean stress value. From tables 1 and 2 it is seen that for a work function of 1·2 ev, the Schottky emission is the larger until the stress reaches a value greater than 2 MV cm^{-1}. This neglects the addition to the F–N value due to the temperature increase, as well as the changes in the Schottky value caused by the barrier-peak tunnelling component and by the transmission coefficient calculated in Appendix B. For average stresses of about 1 MV cm^{-1} it would appear that for realistic values of field intensification, the Schottky type processes would still be the dominant ones.

Real surfaces are usually multicrystalline or have patches of adsorbed materal on them and these result in small areas of differing work functions. The resulting differences in potential between these areas cause patch fields, which ensure that the electron energy change on crossing the barrier is constant and is not determined by the immediate properties of the area from which it leaves. These patch fields may be detected by the increases which they cause in the slope of the Schottky emission plot when the applied field is in the range of their values. For vacuum thermionic emission this range is normally below 5 kv cm^{-1} but since the highest stress at which they have been reported (25 kv cm^{-1}) (Dolan and Dyke 1954) corresponds to extremely small areas of non-uniformity, it is unlikely that this is the cause of anomalous Schottky slopes appearing in high field liquid conductivity experiments.

The presence of a liquid at the surface of a metal influences the work done to emit an electron from the metal to beyond the range of the image force. The change in the value of the vacuum work function has two contributions :

(i) There may be a change in the electrical double layer at the surface. This may be due to a change of the inherent double layer of the surface and to the electrical properties of adsorbed molecules.

(ii) Since the final energy state of the electron will be the energy of its immersion in the liquid, the work function will be reduced by the electron immersion energy.

If the liquid is assumed to behave as a classical dielectric of permittivity ϵ_r within the same range of validity as that of the image law ($x > x' \simeq 3$ Å), an estimate of the order of the immersion energy is given by $e^2(1-1/\epsilon_r)/4x'$.

For hexane $\epsilon_r = 1{\cdot}89$ and this energy reduction is 0·56 ev. This value is critically dependent upon the assumed value of x'.

§ 3. The Conduction Particle

Although the present experiments show that the conduction current over a large stress range can be explained as electron emission from the cathode in fair agreement with emission theory, no definite experimental evidence of the nature of the conduction particle, within the liquid under these high field conditions, has yet been given. Satō, Nagao and Toriyama (1956) have recently given evidence of space charge limited ionic conductivity, at stresses up to 5 kv cm^{-1}, in insulating oil and the values they obtain for the mobility are in good agreement with those given for low stresses in the liquid hydrocarbons by Adamczewski (1937). In a theory of electrical breakdown Goodwin and Macfadyen (1953) assumed ionic conduction occurred at all stresses up to the breakdown value, because the measurement of the time to the first breakdown in hexane, on each fresh electrode area presented, was almost equal to the expected time for an ion to cross the gap. This experimental evidence has since been questioned by Saxe and Lewis (1955), who showed that the time to breakdown, after the application of voltage, could be greatly increased by spark conditioning the electrodes.

Near breakdown it is probable that the conducting particles are electrons and that these are responsible for the ionization suggested in the present experiments. Crowe, Bragg and Sharbaugh (1954) have indeed assumed that this is so and have suggested that the electronic breakdown mechanism resembles those put forward by von Hippel (1937) and Fröhlich (1940) for crystals. Following this, Lewis (1956) has given a qualitative estimate of the breakdown strength in which he assumes that the electron energy-dissipating vibrations are not those between molecules, but those of the molecular bonds themselves.

If the emitted electrons do appear in the liquid as ions, it is probable that they are unstable in high fields, but this remains to be shown experimentally.

§ 4. Apparatus and Experimental Method

A block diagram of the whole of the apparatus is given in figure 2.

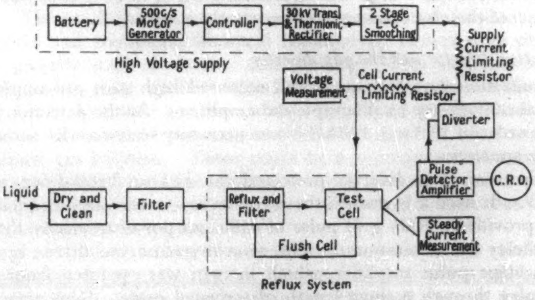

Battery
500c/s Motor Generator
Controller
30 kV Trans. & Thermionic Rectifier
2 Stage L–C Smoothing
High Voltage Supply
Supply Current Limiting Resistor
Voltage Measurement
Cell Current Limiting Resistor
Diverter
Pulse Detector Amplifier
C.R.O.
Liquid
Dry and Clean
Filter
Reflux and Filter
Test Cell
Steady Current Measurement
Flush Cell
Reflux System

Figure 2. Block diagram of apparatus

Liquid preparation

The liquid was stored over sodium and silica gel and when required for use it was passed through a sintered glass filter, of about 1 micron porosity, into the liquid flask of the distillation system. From the flask the liquid was refluxed slowly (Martin 1920), at reduced pressure, on to a further sintered glass filter before passing through to the test cell. Since it was not possible to exclude small dust particles when the test cell was assembled, the liquid was left in the test cell for a short time and then violently flushed back into the liquid flask to remove them.

The gap between the 1 cm diameter spherical electrodes in the test cell was adjusted by a micrometer screw gauge and the axis was horizontal to allow the sphere surfaces to be washed by liquid flushed from the heating flask. The cell had no ' sump like ' pockets and so was self-cleaning on being emptied.

Power supply and current measurement

A 30 kv d.c. supply was obtained by rectifying a 500 c/s alternating voltage. The long time stability of the output voltage was determined by the supply generator, but the changes were quite small and so slow that they caused negligible displacement currents to flow through the test cell. The output ripple voltage was not greater than $\frac{1}{2}$ v mA^{-1} load current and since the maximum total test and measurement current was about 50 μA, the ripple had no ill effect on the operation of the current measuring thermionic electrometer. The high voltage supply was controlled by an auto-transformer controller (House 1954) and the voltage was measured by an electrostatic voltmeter and by a resistor chain and microammeter.

The large range of currents was measured by using several high value resistors (10^6, 10^8, 10^{10} and 10^{12} ohm Morganite resistors), and a thermionic electrometer bridge amplifier. The thermionic valve and the resistors were mounted in a head unit close to the measurement and guard electrodes in the test cell and this was connected to the rest of the amplifier by a screened lead. The resistor and valve head were constructed to allow resistor changeover to take place without the need to switch off the high voltage supply (House and Morant 1954). A galvanometer was used to measure the electrometer output and its sensitivity was made variable to allow the currents to be checked on two resistor ranges. Since the electrometer valve was overloaded when large pulses occurred, the grid current of the electrometer was frequently checked.

The ionization detector and the gap diverter

The ionization detector was a well screened high gain pre-amplifier and a commercial cathode-ray oscillograph and amplifier. As the detection sensitivity was of the order of 10^{-9} A in 100 kΩ it was necessary to use a wire screening cage around the apparatus.

The electronic gap diverter protected the gap on breakdown at voltages up to 14 kv. It used a hydrogen thyratron as a diverter valve and a small neon thyratron provided a 500 v grid pulse of sufficient power to ensure that the grid operation delay was a minimum. The neon thyratron was driven by a conventional two stage pulse amplifier, which in turn was operated from a cathode follower valve through a short length of screened cable. This screened cable was necessary to prevent operation of the diverter by stray voltages induced on long test cell connecting leads. An identical diverter to the one used in the experiments has recently been built and this has a measured operation time of 0·6 to 0·7 μsec. Since such a short time limits the breakdown energy flow into the gap to a value less than its stored energy, there is little advantage in building faster diverters. The minimum tripping voltage was of the order of 10 mv. During the time of current measurement and whilst the ionization detector was being used, the diverter input was connected to the guard electrode which was then grounded through a 330 kΩ resistor.

§ 5. PRESENTATION AND DISCUSSION OF EXPERIMENTAL RESULTS

The earliest tests were made with very clean chromium plated electrodes and the conductivity values were obtained without any conditioning. As explained in § 2.1, the results are most usefully represented on a Schottky plot, since the conduction process may be easily studied by its aid. The individual curves, a representative one of which is shown in figure 3, were not reproducible and the cusps were most probably due to electrode surface changes. This type of curve was associated with violent current pulse spasms.

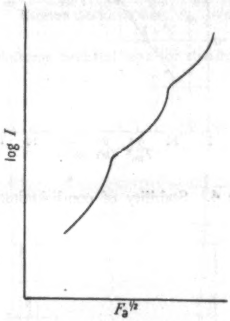


Figure 3. Typical Schottky plot for tests with no conditioning.

After fully conditioning, comparatively smooth and repeatable curves were obtained. The conditioning was only necessary once and, for electrodes of the same material and treatment, identical conductivity curves were obtained for the other polarity with no further conditioning. This suggests that surface conditioning takes place on both electrodes simultaneously, or that ionizable matter of lower dielectric constant within the liquid moves from the high field regions. The former may be a surface ion migration effect and the latter due to minute gas bubbles. These could be a consequence of using spherical electrodes and probably explains why conditioning has been found to be extremely difficult for parallel plane electrodes. It is of interest to note that Green used a sphere–plane electrode system.

The use of spherical electrodes made the analysis of the results difficult, so that, following the remarks in § 2.1, the experimental curves will be compared with the calculated curves of the Schottky emission current from spherical electrodes.

To determine the stability of the conditioned state, a test was made using chromium electrodes after conditioning at a stress greater than 1·2 MV cm^{-1}. The initial curve (see figure 4) has been plotted together with a set of results taken after the test cell had been left open to the atmosphere, through a 5 to 10μ porosity sintered glass filter, for 24 hours. A further curve in the same figure shows the results taken a short time later, after partial conditioning up to a stress of about 1·1 MV cm^{-1}.

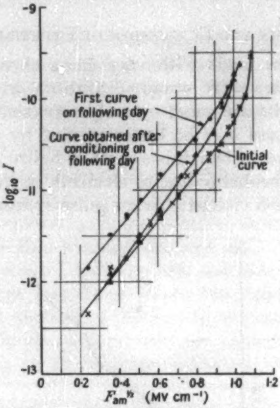


Figure 4. Stability of conditioning test.

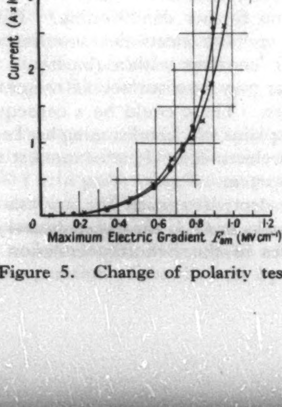


Figure 5. Change of polarity test.

An attempt was made to determine if the current depended upon the cathode metal, by taking current measurements for cathodes of different materials and using the same anode, but since the results even for electrodes of similar metals could not be accurately repeated after opening the cell, a polarity reversal technique had to be used. Figure 5 shows the two curves obtained for a pair of chromium electrodes, whilst figure 6 shows the corresponding curves for a silver and a chromium electrode. The initial curve was rechecked in each case. Whilst the method of preparation for all the electrodes was the same, the silver electrodes may not have had such a good surface polish as the chromium ones, and this may have caused part of the difference between the currents. The Schottky plots for the mixed electrodes were slightly curved.

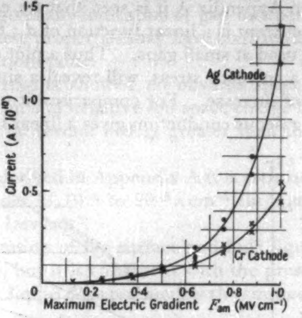


Figure 6. Polarity reversal test for dissimilar electrodes.

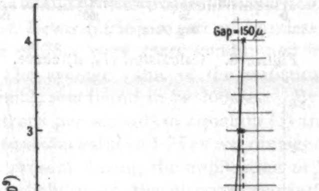


Figure 7. Experimental current–stress curves.

In order to detect ionization processes in the bulk of the liquid a series of conductivity tests were made for a number of distances between similar electrodes. The results of these tests are shown in figure 7. All these results, except the ones for gap distances of 75 and 150μ, were repeated by resetting the gap and measuring the currents again. The liquid broke down at a stress of 1·55 MV cm^{-1} just after checking the results for a gap distance of 25μ.

From equation (2) in Appendix A it is seen that for each value of maximum stress F_{am}, the emitted current is a linear function of d ; this is a consequence of the spherical electrodes used at small gaps. Thus a plot of the current I against the gap distance d, for a constant stress, will reveal a simple ionization process by an exponential type of increase. For comparison it may be recalled that the Townsend α process of gaseous conduction gives a linear increase of the (log I, d)

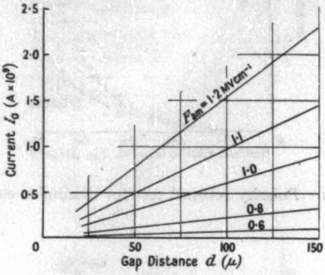


Figure 8. Calculated (I_0, d) curve.

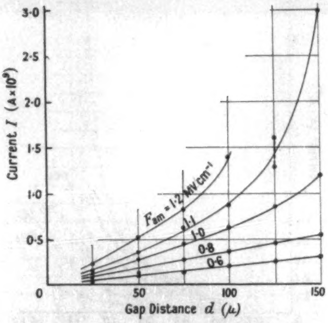


Figure 9. Experimental (I, d) curve.

curve for currents in a parallel plane gap. A plot of the values calculated from equation (3), for $e\phi = 1{\cdot}2$ ev is given in figure 8 and similar plots of the experimental values, given in figure 7, are shown in figure 9. It will be noted that since the curves are ones for which $F_{\mathrm{am}} = \text{const.}$, deductions from them are not limited by the condition of Schottky emission. The only limitation is that the emission law must increase at a fast rate with increase of stress, to allow the greater proportion of the current to flow in a region within the range of validity of equation (1) in

Appendix A. This criterion is satisfied with the usual emission laws. A comparison of figures 8 and 9 shows that there is a significant curving of the experimental curves for the higher stresses and suggests that bulk ionization processes occur within the liquid at these stresses and in addition that there may be a changing value of the effective work function which is dependent upon the magnitude of the current flowing and the stress F_{am}.

None of the plots gives any indication of gap independent liquid conduction properties, although these properties may have contributed to the experimentally determined work function value.

The experimental results showed no obvious space charge effects and the estimation of the electron and negative ion space charge limited values of current density show them to be several orders greater than the experimentally determined currents.

Using the relation derived in Appendix A it is estimated that the true current densities are of the order of 10^{-8} to 10^{-9} A cm^{-2} for a stress of 0·5 MV cm^{-1} and 10^{-7} to 10^{-8} A cm^{-2} for 1 MV cm^{-1}.

A satisfactory separation of the surface and bulk liquid effects in conduction has not yet been made, but it is consistent with the present experimental results to assume that the bulk liquid does not modify the emitted current until ionization begins to occur.

Since little is known of the high field emission of electrons from metal surfaces into liquids, it was thought that a knowledge of the electrode emission in air under similar conditions would provide a useful comparison. Using the apparatus described by Saxe and Lewis (1955), a pair of satinless steel electrodes, 1 cm diameter and at a gap of 75μ, were spark conditioned in air, with an efficient electronic diverter. The average value of the statistical time lag in air, t_{av}, at a stress of 1160 kv cm^{-1}, was found to be 260μsec. By equating the emission current ($I_0 = \mathbf{e}/t_{\mathrm{av}}$ electrons per second) to equation (3) and putting $\epsilon_r = 1$, $\overline{D} = 1$ and $T = 300$°K, a work function value of 1·75 ev was obtained. This value assumes that the first electron present during the application of the pulse would cause breakdown. If the probability of the electron causing breakdown were less than unity, then the work function value would be lower than 1·75 ev.

§ 6. SUMMARY OF EXPERIMENTAL RESULTS AND CONCLUSIONS

It would appear that conditioning is due to the removal of minute gas bubbles from the high stress region of the liquid and from the central portion of the electrode surface, or due to the stripping of lightly adsorbed layers from the electrode surface. At the higher stresses, when stable conditions are reached, the more tenaciously adsorbed layers of material must still be present. Hancox (1956) working in this laboratory on the high voltage impulse strength of oil, has recently shown that, by controlled exposure of cleaned electrodes to the atmosphere, repeatable impulse breakdown strengths ranging from 1000 to 2000 kv cm^{-1} may be obtained. This is in agreement with the present work since the impulse breakdown strength under these conditions would be expected to be a measure of the electron availability. It must be borne in mind that a small change in the work function gives rise to large changes in the current emitted.

The less tenaciously adsorbed layers might be expected to be in quasi-dynamic equilibrium with the liquid or liquid impurities and thus give the current–temperature hysteresis effect reported by Le Page and Du Bridge (1940). That adsorbed gases give rise to pulses was noted by Baker and Boltz (1937). It is believed that in the present work the liquid purity was high and allowed the conditioned state to be held for long periods. The conditioning action may be helped by a non-uniform field geometry and conditioning is only necessary for one polarity. By using a pair of electrodes of silver and chromium in the same cell and measuring currents using each in turn as a cathode, repeatable conduction curves show that the current values were in the ratio of only 3 : 1 at an applied stress of 1 MV cm^{-1}. Check tests on pairs of similar electrodes gave identical curves within the range of experimental error. Thus there would appear to be no merit in trying to correlate these results with vacuum work function values.

In contrast to other conductivity results in liquids, the author obtained Schottky (log I, $(F_a)^{1/2}$) plots of the conduction curves which, for the well conditioned electrodes, were straight lines over a large range of stresses. The calculations in Appendix A and elsewhere show that, for the experimental electrodes chosen, the Schottky plot was still valid, but was slightly altered in slope. The values over many tests varied between $2{\cdot}5 \times 10^{-3}$ and 4×10^{-3} in units of $\log_{10} I$ and v cm^{-1}.

The work function value calculated at 1000 kv cm^{-1} for each experimental curve was not greater than 0·03 ev higher than that for 500 kv cm^{-1}, but throughout the tests the average values varied between 1·04 and 1·13 ev. This work function value may possibly include a bulk liquid component independent of gap. These values compare with the limiting value of 1·75 ev obtained for the spark conditioned electrodes tested in air.

Schottky's formula shows that the Schottky emission current, from spherical electrodes at small gaps, is proportional to the gap distance (see figure 8). This linear dependence is maintained for other electron emission processes, provided only that their emitted current–stress dependence increases at least as rapidly as the Schottky one. The derived experimental (I, d) curve of figure 9 shows this linear dependence up to a stress of 800 kv cm^{-1}, but at stresses of 1 MV cm^{-1} or greater the slope of the current curve increases rapidly at the larger gap values. This increase in current dependent on gap could indicate an ionization process similar to that in Townsend's classical theory for gases. Unfortunately the spherical electrode geometry makes it difficult to determine the value of α in the simple ionization expression $I = I_0 \exp(\alpha d)$. It seems less likely that the gap-dependent current increase may in part be due to a dependence of the work function upon the current, but this possibility cannot be eliminated. Experimental proof would be required to substantiate these ideas.

No space charge effects were noted and calculations of the values of the limiting currents for ion and electron space charge conditions showed them to be many orders greater than the currents experimentally determined.

ACKNOWLEDGMENTS

The experimental work was carried out at Queen Mary College during the tenure of the Henley Research Fellowship which is gratefully acknowledged. The author thanks Dr. H. Tropper for his encouragement and guidance, Professors W. J. John and M. W. Humphrey Davies for the facilities made available, and colleagues for their helpful discussions. He is also indebted to Mr. H. Astbury, Chief Engineer of Messrs Crompton Parkinson Co. Ltd., Chelmsford, for the manufacture of equipment.

APPENDIX A

AN ESTIMATE OF THE SCHOTTKY EMISSION CURRENT FLOWING BETWEEN TWO SPHERICAL ELECTRODES

In the experimental results (see, for example, figure 5) it is seen that, at the higher stresses, the conductivity current increases rapidly with increase of stress. This suggests that a relatively small area may be responsible for most of the emission current. Within this small area and for very small gap distances, the electrical stress on a small annular ring of radius $R \sin \theta$ (see figure 10)

$$F_a \simeq F_{am}/\{1-2R(1+\cos\theta)/d\} \qquad \ldots\ldots(1)$$

where F_{am} is the maximum surface stress in the gap. That this is a good approximation to the correct surface stress has been shown by the author.

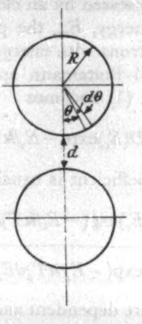


Figure 10. Spherical electrodes.

The contribution to the total current from the annular ring is

$$\delta I_0 = 2\pi R^2 \delta(\cos\theta) j(F_a)$$

where $j(F_a)$ is the emission current density into a medium of dielectric constant ϵ_r Using equation (1) it can be shown that

$$I_0 = -\pi R d \int_{F_{am}}^{F_{amin}} j(F_a) F_{am}/(F_a{}^2) dF_a \qquad \ldots\ldots(2)$$

and using the Schottky expression for emission current density

$$I_0 = \pi R A \overline{D} T^2 d \exp(-\boldsymbol{e}\phi/\boldsymbol{k}T) \int_{F_{amin}}^{F_{am}} \{F_{am}/F_a{}^2\} \exp\{\boldsymbol{e}(\boldsymbol{e}F_a/\epsilon_r)^{1/2}/\boldsymbol{k}T\} dF_a. \qquad \ldots\ldots(3)$$

This integral has been graphically evaluated, within its range of validity, for many values of F_{am} and using $T = 300°\text{K}$ and $\epsilon_r = 1{\cdot}875$. These calculations show that the contribution decreases rapidly as the surface stress is reduced on moving away from the axis. The value of the integrand at $F_a = 0{\cdot}7\,\text{MV cm}^{-1}$ is only 0·4 of its value for $F_{am} = 1{\cdot}4\,\text{MV cm}^{-1}$ and this occurs for $\theta = 10°$, showing that the approximations in equations (1) and (2) are permissible. It will be seen that the integrand does not represent the required function for small values of F_a ; the graphical function used tended to zero as $F_a \rightarrow 0$.

On plotting these results in the $(\log_{10} I,\ F_{am}{}^{1/2})$ form the Schottky slope was found to be reduced from its parallel plane electrode value of $4{\cdot}82 \times 10^{-3}$ to

$4{\cdot}2 \times 10^{-3}$ in units of $\log_{10} I$ and v cm^{-1}, both values being corrected for emission into hexane ($\epsilon_r = 1{\cdot}875$). The intercept of the plot could not be used to determine the work function, so it was calculated for the stress values within the range of validity of the equation (3).

APPENDIX B

It can be shown that the Schottky emission current density is given by the expression

$$j(F_a) = \boldsymbol{e}(2\boldsymbol{m}/\boldsymbol{h}^2)^2 \pi \boldsymbol{h}/\boldsymbol{m} \int_0^\infty dE_r \int_{E_{zm}}^\infty \frac{D(E_z)dE_z}{\exp\{(E_z+E_r-E')/\boldsymbol{k}T\}+1} \quad \ldots\ldots (1)$$

where E_z is the energy possessed by an electron due to movement perpendicular to the surface, E_r the energy possessed by an electron due to movement parallel to the surface, E' the Fermi energy, E_{zm} the peak barrier energy, $D(E_z)$ the transmission coefficient for electrons with energy E_z.

Since $E_{zm} > E'$ the Maxwell–Boltzmann approximation can be used and the second integral in equation (1) becomes

$$\exp\{-(E_r-E')/\boldsymbol{k}T\}\int_{E_{zm}}^\infty D(E_z)\exp(-E_z/\boldsymbol{k}T)dE_z.$$

The average transmission coefficient is usually defined as

$$\bar{D} = \frac{\int_{E_{zm}}^\infty D(E_z)\exp(-E_z/\boldsymbol{k}T)dE_z}{\int_{E_{zm}}^\infty \exp(-E_z/\boldsymbol{k}T)dE_z}. \quad \ldots\ldots (2)$$

It is seen that $\bar{D}$ is temperature dependent and since, for the Schottky barrier, $D(E_z)$ is dependent on field, $\bar{D}$ is dependent on both field and temperature.

Values of $D(E_z)$ have been given by Guth and Mullin (1941) for the applied field values of 40, 122·5 and 250 kv cm^{-1}. Using these values, $\bar{D}$ has been graphically evaluated for values of $T = 300°$ K (normal temperature) and 2400° K and the results are shown in table 3.

Table 3. Transmission Coefficients for Schottky Emission

F_a (kv cm^{-1})	40	123	250
$\bar{D}$ (T=300°K)	0·85	0·815	0·62
$\bar{D}$ (T=2400°K)	0·967	0·962	0·93

These values are for a Schottky metal–vacuum interface, but the metal–liquid interface might be expected to give similar results. It would be interesting to know the values of $\bar{D}$ at higher stresses, but extrapolation would be unwise and the calculations of Guth and Mullin were exceedingly tedious.

REFERENCES

ADAMCZEWSKI, J., 1937, *Ann. Phys., Paris*, **8**, 309.

BAKER, E. B., and BOLTZ, H. A., 1937, *Phys. Rev.*, **51**, 275.

BURGESS, R. E., KROEMER, H., and HOUSTON, J. M., 1953, *Phys. Rev.*, **90**, 515.

CROWE, R. W., BRAGG, J. K., and SHARBAUGH, A. H., 1954, *J. Appl. Phys.*, **25**, 1480.

DOLAN, W. W., and DUKE, W. P., 1954, *Phys. Rev.*, **95**, 327.

FOWLER, R. H., and NORDHEIM, L., 1928, *Proc. Roy. Soc.* A, **119**, 173.

FRÖHLICH, H., 1940, *Rep. Progr. Phys.*, **6**, 411 (London: Physical Society).

GOODWIN, D. W., and MACFADYEN, K. A., 1953, *Proc. Phys. Soc.* B, **66,** 85.
GREEN, W. B., 1956, *J. Appl. Phys.*, **26,** 1257.
GUTH, E., and MULLIN, C. J., 1941, *Phys. Rev.*, **59,** 578; 1942, *Ibid.* **61,** 339.
HANCOX, R., 1956, *Nature, Lond.* **178,** 1305.
VON HIPPEL, A., 1937, *J. Appl. Phys.*, **8,** 815.
HOUSE, H., 1954, *J. Sci. Instrum.*, **31,** 261.
HOUSE, H., and MORANT, M. J., 1954, *J. Sci. Instrum.*, **31,** 342.
LEPAGE, W. R., and DU BRIDGE, L. A., 1940, *Phys. Rev.*, **58,** 61.
LEWIS, T. J., 1956, *J. Appl. Phys.*, **27,** 645.
MARTIN, W. H., 1920, *J. Chem. Phys.*, **24,** 478.
NORDHEIM, L., 1928, *Proc. Roy. Soc.*, A, **121,** 626.
SATŌ, T., NAGAO, S., and TORIYAMA, Y., 1956, *Brit. J., Appl. Phys.*, **7,** 297.
SAXE, R. F., and LEWIS, T. J., 1955, *Brit. J. Appl. Phys.*, **6,** 211.
SOMMERFELD, A., and BETHE, H., 1934, *Handbuch der Physik*, **24,** Part 2, 436.

Contact Effects resulting from Compression and Flash Sintering in Graphite Powders

By F. A. LEWIS†, J. ORR AND A. R. UBBELOHDE
Department of Chemical Engineering, Imperial College

MS. received 13*th March* 1957, *and in final form* 27*th June* 1957

Abstract. Experiments are described on contact resistance effects in graphite powders : (i) as these are progressively compressed up to 3600 kg cm^{-2} in the cold : this produces a marked decrease of electrical resistance, though it seems unlikely that much compression of the actual crystal lattice of graphite is effected ; (ii) when loosely pressed powder is subjected to flash discharges from a bank of charged condensers ('flash sintering') : this is in order to explore possible mechanisms of sintering that avoid any prolonged heating of the graphite at high temperatures.

In the case of powdered polycrystalline graphite, both (i) and (ii) lead to a marked lowering and stabilization of the electrical resistance. Their influence on the thermoelectric power is discussed. 'Flash sintering' promises to be a useful and accessible technique in studying the conductance of crystal compounds of graphite. The results throw a significant light on the nature of the conductance barriers in graphite powders.

§ 1. INTRODUCTION

PREVIOUS work (McDonnell, Pink and Ubbelohde 1951) has shown that the formation of crystal compounds has marked effects on the electronic properties of graphite. Such effects have been interpreted on a model which attributes both electron donor and electron acceptor properties to the amphoteric layers of fused carbon rings. Intercalation between the layers of acceptor groups such as bromine or aluminium chloride (Orr and Ubbelohde, to be published) reduces the number of electrons in a full band in the carbon macromolecule, whereas intercalation of donors such as the alkali metal atoms injects electrons into the empty band lying just above it. Either process greatly increases the electrical conductivity, and also influences other electronic properties, such as the thermoelectric power (McDonnell, Pink and Ubbelohde 1951) and the Hall effect (Currie *et al.* 1956).

In pursuing this line of investigation on the chemical physics of aromatic systems, one major difficulty springs from the circumstance that most methods yet evolved for preparing lamellar crystal compounds of graphite lead to powders. When properties such as the electrical conductance are studied, it is necessary to determine how far these are influenced by contact effects between particles of the powder, and surface and edge effects between individual crystallites in each particle. Interpretation of such contact effects in terms of fundamental electronic processes in disordered crystals is of considerable theoretical as well as practical importance.

---

† Now at Department of Chemistry, Queen's University, Belfast, N. Ireland.

In powder metallurgy the process of inducing true lattice joins between particles of powder is known as sintering ; this may also induce crystal growth. Conventional procedures for sintering particles of graphite would require the use of a high temperature, often exceeding 1500°c. Apart from other objections this makes conventional high temperature sintering inapplicable to crystal compounds, which would decompose at much lower temperatures. The present paper records studies of two possible alternative routes for removing uncertainties arising from contact effects in powdered graphite, with the object of extending suitable techniques to the crystal compounds in due course.

Compression in the cold to high pressures can increase the bulk density of powdered graphite to values approaching and exceeding that of the original polycrystalline block from which the powder was prepared. A parallel increase in electrical conductance is found, as is described below. The extent of the change indicates that electron contact barriers between crystallites of graphite are comparatively unimportant and that increasing the density mainly serves to decrease the effective path of flow of the current. This behaviour contrasts with certain other metal powders in which particles may be coated by skins of oxide which have considerably higher resistance than the metal.

Flash sintering appears to remove certain reactive regions at contact points between the particles, probably by inducing local bond rearrangement leading to welding. This has interesting consequences for the electronic properties of loosely contacted graphite powder and suppresses the drift of these properties with time.

§ 2. EXPERIMENTAL

2.1. *Materials*

For measurements under pressure, a block of polycrystalline graphite kindly provided by the Atomic Energy Research Establishment, Harwell was powdered, and the fines passing a 300 mesh sieve were used. The original cylindrical block graphite had a specific resistance along its extrusion axis of $1 \times 10^{-3}\Omega$ cm and across the extrusion axis of $3 \times 10^{-3}\Omega$ cm. Its apparent density was 2·00 g cm^{-3}, determined by flotation. The bulk density was 1·68 g cm^{-3}. This powder was also used for many of the flash sintering experiments. In addition, raspings from a rod of Acheson graphite were used for some of the measurements of thermoelectrical power.

Air was not excluded during powdering, so that any free valencies produced by disrupting carbon macromolecules would almost certainly oxidize to peroxide, quinone or OH groups.

2.2. *Pressure Measurements*

Previous studies of the effects of compression in the cold on the electrical conductance of powders have not been extensive, and have frequently been limited to particles coated with skins of different chemical composition and considerably lower conductance. When such skins can be avoided, as appears to be the case in the present studies, the effects of compression on the electronic properties are of considerable fundamental interest. There is, however, a technical difficulty in applying high pressure to powders. If a fluid is used to ensure uniform application of pressure on each particle of the powder, films of the fluid can creep between contacts and influence their electrical properties by their own resistance. If dry powders are compressed, non-uniform pressures on individual particles tend to be built up by a system of forces and by friction at the walls of the container used. These forces involve proportionately more serious lack of uniformity if the pressure is exerted on small areas and long lengths of a column of powder, as is desirable with a good conductor in order to attain conveniently measurable electrical resistances. Ways of meeting these difficulties in the case of powdered graphite are described in what follows.

Figure 1 shows the resistance cell used for definitive measurements, evolved after many exploratory trials on powdered graphite. Measurement of electrical resistance and apparent density could be carried out simultaneously with this apparatus.

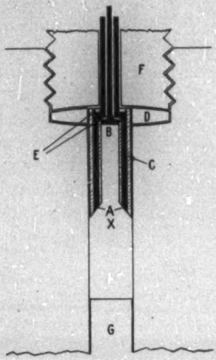


Figure 1.

The cell into which the powder was compressed was a close fit in the top of a high pressure cylinder. The lower end A of a copper cylinder formed one electrode of the cell. The top electrode B was a disc at the end of a copper rod. The electrodes were insulated from one another and from the steel cylinder walls by thin tubes of ethoxyline resin (silica filled). The outer insulating tube C fitted tightly into the lens ring D which helped to align the cell. Both electrodes were shouldered so that the high pressure thrust was transmitted, through insulating washers E, also made from silica filled ethoxyline resin, to the screwed plug F. Above these shoulders the electrodes consisted of a coaxial cylinder and rod. Soldered connections were made to these above the plug F after the cell was in position. Electrical resistance was measured directly by a Smith's low resistance bridge. In order to vary the column length of powder, the maximum length of the cell (5·0 cm) could be reduced, by screwing further short copper cylinders of the same diameter (0·75 cm) into the top electrode.

The cell constant and the resistance of the leads were first measured using salt solutions and mercury. Any changes of cell dimensions before and after application of pressure were measured by the cathetometer and by recalibration of the cell constant with pressures below 3000 kg cm^{-2}. Any irreversible changes in cell dimensions led to errors in the electrical resistance measurements of less than 1%. Corrections for reversible volume changes of the cell materials were of the same magnitude. Errors of this order do not affect the conclusions discussed below.

The powder under investigation was pressed into the cell by a piston G moving in the high pressure cylinder. A vernier on this piston moved against a graduated scale and was viewed by a remote optical system. The volume occupied by a known mass of graphite (usually about 15 g) could thus be calculated. For each measurement the given pressure was alternately released and reapplied until the electrical resistance had reached a limiting value indicated by no change in resistance on further application of pressure. This operation is analogous to that of tamping. By this means greater reproducibility of results is ensured. Moreover, if the pressure was steadily increased in one single thrust without tamping, damage to the lower electrode and insulating sleeves of the cell resulted due to the powder becoming densely packed at X (figure 1) before the space between the electrodes was filled uniformly.

Because of the uncertainties in the packing of the powder, the usual plot of resistance against column length to test for any appreciable contact resistance between the copper electrodes and the column of powder (McDonnell, Pink and Ubbelohde 1951) cannot be very precise. However, the ratio of the resistance to length of column was found to be practically independent of column length, indicating that any contact resistance between copper and graphite powder could be neglected in the present experiments.

In the course of cycles of observations on electrical resistance and on pressure, which were measured simultaneously, it was noted that as the total compression increased fewer tamping thrusts were necessary before the electrical resistance at the given pressure reached its limiting value. Values of the electrical resistance for unit length were reproducible to within 5% at any given pressure, for different samples of the same powder. Hysteresis was observed in both volume and resistance readings when the pressure was released.

2.3. *Results for Powders under Pressure*

Figure 2 (*a*) records a typical run of pressure measurements. Volume changes were recorded simultaneously. Volume changes at higher pressure (which

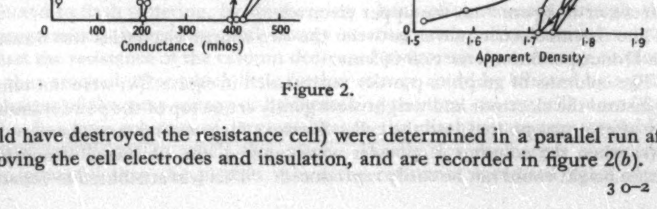
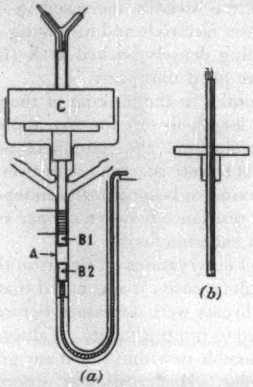


Figure 2.

would have destroyed the resistance cell) were determined in a parallel run after removing the cell electrodes and insulation, and are recorded in figure 2(*b*).

2.3. *Flash Sintering Experiments*

Figure 3(*a*) illustrates the apparatus used for measurements on the electrical resistance and on the thermo electric voltage of columns of powdered graphite. The material was packed in a Pyrex tube A and was maintained between two cylindrical copper electrodes B1, B2, of cross sectional area 0·085 cm^2. These electrodes were a sliding fit in the Pyrex tube and were faced with thin platinum sheet, hard soldered to the copper, so that the barrier contact was between bright platinum and graphite. Each of these copper electrodes formed one arm of a copper–graphite thermocouple. To measure its temperature by means of a copper–eureka thermocouple the top electrode carried the eureka wire down its centre and was insulated from it by a glass capillary. A section through this electrode is shown in figure 3(*b*). For resistance measurements, most of the apparatus could be immersed in a bath with thermostat control to ±0·05° c and no heating coil was used. For special purposes, the sample in the Pyrex tube could be maintained under oxygen-free nitrogen, but this was not found necessary in the experiments described below.



(*a*) (*b*)

Figure 3.

To measure thermoelectric power, a heating coil of 'brightray' wire was wound on the outside of the glass cell in the vicinity of the top electrode so that a controlled temperature difference could be maintained between the two electrodes. The powder column itself and the lower electrode were at the temperature of the thermostat bath in which the cell was immersed. A lead weight C exerted a pressure of 13 kg cm^{-2} on the upper electrode.

The thermoelectric power between the two copper electrodes was measured on a Doran potentiometer to 0·005 mv.

The columns of graphite powder represented in figure 3(*a*) were not tamped but instead the electrode and weight were gently set on top of the powder immediately after it was poured into the cell. Reasons for avoiding tamping were :

(i) when the columns of powder were tamped, the resistance of any given column height could not be easily reproduced. This was attributed to variations in the degree of tamping and possibly in the preferred orientation of particles; these were difficult to eliminate. There was also a prolonged drift of resistance with time after tamping, which made measurement very inconvenient. The resistance decreased progressively, suggesting slip followed by welding at points of contact between particles.

(ii) By using untamped columns the effect of flash sintering on poor contact points of high resistance could be more readily followed. Figure 4(*a*) records measurements of resistance plotted against column length for untamped powders. A curve is obtained instead of the straight line plot normally expected for uniformly packed columns. As the column height increased the packing became progressively less close, and the effective resistance per unit length increased.

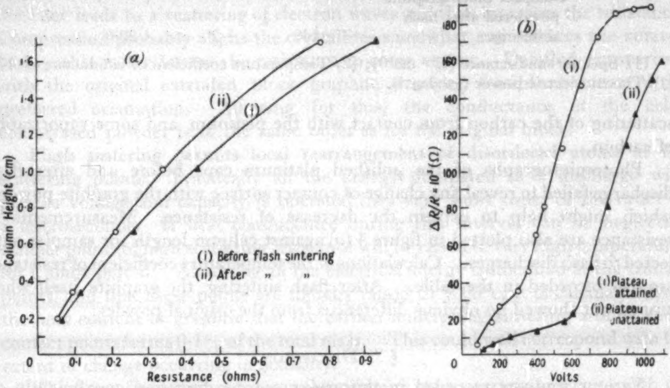


Figure 4.

With the packing precautions as described, quite steady readings of resistance were, however, obtained. After obtaining the highest resistances and column lengths recorded in figure 4(*a*) the temperature coefficient of total resistance was determined for the longest columns, to minimize any leads effects. Thermoelectric measurements were also made at various bath temperatures. A plot of thermo-e.m.f. against temperature interval between the two carbon–copper junctions gave a straight line (not shown). From its slope the thermoelectric power was obtained as recorded in the table.

2.4. *Effects of Flash Sintering*

The column lengths of graphite powder (figure 4(*a*)) packed as above were subjected to flash sintering, by discharging a bank of condensers through them of progressively increasing total energy starting with various initial voltages. At first the resistance of the column decreased progressively. A stage or plateau was then reached, after which little further change could be effected by repeated discharges of the same energy or by increasing the total energy still more (figure 4(*b*). For long columns of large resistance the discharge energy available to us was insufficient to reach this stage. Finally, for discharges considerably more powerful than the plateau values the resistance increased again, due to

Electronic Properties of Graphite before and after Flash Sintering

Sample (A.E.R.E.)	(1)	(2)	(3)
Compact block graphite parallel to extrusion axis.	1000	1·2	−0·1
Compact block graphite perpendicular to extrusion axis.	333	1·2	+1·9
Compact block graphite powdered before flash sintering.	62·5	1·4	+23·0
Compact block graphite powdered after flash sintering.	140	1·2	+20·0

(1) Specific conductance (Ω^{-1} cm^{-1}); (2) Temperature coefficient of conductance $\times 10^3$; (3) Thermoelectric power (μv deg^{-1}).

scattering of the carbon from contact with the platinum, and some vaporization of carbon.

Photomicrographs of the polished platinum caps before and after flash discharge failed to reveal any change of contact surface with the graphite powder, which might help to explain the decrease of resistance. Measurements of resistance are also plotted in figure 4 (*a*) against column length for samples subjected to flash discharges. Calculations of the temperature coefficient of resistance are also recorded in the table. After flash sintering, the graphite itself when tipped out showed no obvious differences from the original powder.

§ 3. DISCUSSION

Various findings recorded in this paper can be discussed conveniently on the basis of the contacts between edges and corners of small 'perfect' crystallites of graphite. In the particles of powder these may be only a little thicker than macro-aromatic molecules of fused carbon rings forming each layer of the perfect crystallite. Various chemical differences may be expected at the edges of such crystallites or macro-aromatic molecules.

(i) Free carbon valencies may be bonded to oxygen atoms. An analogy is when hexaphenylethane interacts with atmospheric oxygen. Bonded oxygen may be present as peroxide bridges or quinone groups. There may also be OH groups if the hydrogen which is often present in graphite has not been completely expelled during its original baking, or if water is absorbed chemically from the atmosphere when the polycrystalline lump is ground and sieved.

(ii) Neighbouring macro-molecules may be jointed by actual carbon bonds, probably under considerable strain, again as in hexaphenylethane and similar molecules. More than one bond may join macro-molecules of carbon in favourable cases.

The effects of cold pressing on the bulk density of graphite powders suggests that random stacking of the crystallites in the original powdered sample appears to be modified to give a very considerable degree of alignment at high pressures. Figure 2 (*b*) indicates that the maximum apparent bulk density achieved in the cold was higher than that (1·68 g cm^{-3}) of the block with preferred orientation from which the powder was prepared. Considerable strain energy appears to be stored in the compressed powder, which expands again on releasing the pressure. Detailed examination of the hysteresis curves in figure 2 indicated that actual compression of the crystal lattice is unlikely at the pressures achieved. This concurs with estimates of the compressibility of graphite (Riley 1945, Basset 1941).

The effects of cold pressing on the resistance and volume occupancy of graphite powders can be summed up by the conclusion that any electron barriers at contact points between neighbouring crystallites in a particle, or between particles, are only small, and are much less, for example, than between particles of metal where skins of oxide intervene at contact points. Though the electron barriers are only small at contacts between graphite particles, the crystal disorder leads to a scattering of electron waves and thus increases the resistance. Compression probably aligns the crystallites somewhat and reduces the current path ; this factor lowers the resistance to some extent. Detailed comparison with the original extruded block graphite was not feasible because of its preferred orientation. Allowing for this, the conductance of the cold-compressed powder is of the same order as for the original block.

Flash sintering permits local rearrangement of disordered atoms at the scattering points. Practically all the energy dissipated in a discharge with plateau voltage and capacity is liberated in a very short time, of the order of 5 microseconds. If heat conductance during this interval can be neglected, contact points of high scattering resistance will be raised to very high temperatures for a short period. Assuming all the electrical energy is liberated at the contact points, and that these points are thereby raised to 4000°c, it is calculated from the heat content of graphite that the carbon undergoing sintering around these contact points forms 0·1% of the total mass. This could well correspond with the extent of change occurring in actuality.

The flow of graphite powder under pressure shows certain analogies with the discharge process, since mechanical strain energy is probably released by sudden slipping at contact points, which thereby become momentary hot spots. The drift of resistance in tamped columns under light pressures can in fact be compared with a progressive chemical polymerization, in which the macro-molecules of graphite weld together at the hot spots each time a slip occurs. This corresponds with the melting and solidification in other frictional processes ; the high melting point of graphite around 4000°c permits hot spots to attain exceptionally high temperatures.

These conclusions about the contact in graphite powders can be further elucidated from a consideration of the aromatic character of the macro-molecules in graphite. At the edges these cannot hold more than one or two impurity groups such as $=\mathrm{O}$, $-\mathrm{OH}$, etc. By analogy with the behaviour of smaller aromatic molecules, such groups will form part of the aromatic ensemble of the macro-molecule, by electronic conjugation. Unlike powdered metals, powdered graphite cannot therefore readily form thick barrier layers of low conductance, separating individual crystallites. This conclusion is highly important in developing the study of crystal compounds of graphite obtainable only in powder form.

The effects of flash sintering on the electronic properties give additional and more detailed information, particularly in the case of electronic properties of the macro-molecules which are more sensitive to edge effects than the overall conductance. In a model of graphite discussed by Mrozowski (1952) it is pointed out that valency defects at specific carbon atoms in graphite remove corresponding electrons from the full band of the perfect crystal, to lower energy levels. Such valency defects if not too numerous increase the conductance, and affect other electronic properties of graphite correspondingly.

The present experiments are primarily concerned with edge defects. The first point to note is that whilst the thermoelectric power of the original block graphite against copper lay between $-0{\cdot}1$ and $+1{\cdot}9\mu$v deg^{-1} according to the orientation, the same material in powdered form in a lightly packed column shows a thermo electric power of $+23{\cdot}0\mu$v deg^{-1}. This comparatively large change would be expected if a large increase in edge defects and possible edge oxides occurs on powdering, thus depleting the band in the graphite crystallites.

Because of the increased conductance path in the powder, direct comparisons of the conductance in block and powder do not give immediate information about changes in the Π bands. However, differences between the temperature coefficients of conductance for block and powdered graphite again indicate a greater depletion of the Π band in the powders.

The length of column was not found to alter appreciably after flash sintering. But the effective conductance paths are probably shortened by the elimination of electron scattering at contacts between impurity groups and by the direct formation of carbon–carbon bonds. Flash sintering removes a considerable proportion of defect sites, as is confirmed by the changes in electronic properties in the table. The temperature coefficient of resistance drops to the value for block graphite and the thermoelectric power becomes less positive, as defects which deplete the Π band decrease in number.

ACKNOWLEDGMENTS

The authors wish to thank the referees for drawing their attention to the experiments of Bowden and Williamson (1954), with contacts of gold. These would not form thick oxide films separating the metal conductors, and it is interesting that the findings of these authors lend broad support to our present conclusions for graphite particles.

Thanks are due to the Central Electricity Board for their support of this research.

REFERENCES

BASSET, J., 1941, *C. R. Acad. Sci., Paris*, **213**, 829.

BOWDEN, F. P., and WILLIAMSON J. B. P., 1954, *Research Correspondence*, Vol. 7, Number 10, October 1954.

CURRIE, L. M., HAMISTER, V. C., and MACPHERSON, H. G., 1956, *Proceedings of the International Conference on the Peaceful Uses of Atomic Energy* (New York : United Nations) Vol. 8, p. 451.

MCDONNELL, F. R. M., PINK, R. C., and UBBELOHDE, A. R., 1951, *J. Chem. Soc.*, 191.

MROZOWSKI, S., 1952, *Phys. Rev.*, **85**, 609, **86**, 1056.

RILEY, D. P., 1945, *Proc. Phys. Soc.*, **57**, 486.

The Dielectric Properties of Methyl and Ethyl Alcohols in the Wavelength Range 3 cm – 52 cm

BY E. H. GRANT
Department of Physics Applied to Medicine,
The Middlesex Hospital Medical School, London

Communicated by J. E. Roberts; MS. received 15*th April* 1957,
and in final form 10*th July* 1957

Abstract. Measurements have been made of the dielectric properties of methyl and ethyl alcohols in the wavelength range 3 cm–52 cm. It is concluded that the dielectric behaviour of both alcohols can be described in terms of a principal dispersion region at centimetre wavelengths together with a subsidiary dispersion region occurring at millimetre wavelengths.

---

§ 1. INTRODUCTION

SINCE the advent of reliable continuous wave sources of decimetric radiation, there appear to have been no investigations at room temperatures in the main parts of the relaxation regions of methyl and ethyl alcohols. As shown by Saxton (1952) there is lack of agreement between results obtained by previous workers. Work at the high frequency ends of the dispersion regions for both alcohols has been carried out by Lane and Saxton (1952) and Poley (1955) and there is also a measurement on ethyl alcohol recorded by Bolton (1948) at a wavelength of 12·6 cm. The static dielectric constants for both liquids have been measured by Mecke and Reuter (1949).

Since the relaxation wavelengths at room temperatures for methyl and ethyl alcohols are about 10 cm and 30 cm respectively, the present work at free-space wavelengths of 3 cm, 10 cm, 17 cm and 52 cm has the advantage of providing measurements near to, and on both sides of, the centre of the dispersion region. Thus the results are of value, partly for their own sake, and partly to check the conclusions of Saxton and Poley which were made on the basis of their measurements at the shorter wavelengths.

Experimental considerations are dealt with in §2, and in the remaining two sections an attempt is made to deduce the nature of the dispersion curves from all the available results. Molecular considerations have not been included, the purpose of this paper being merely to report new measurements in the microwave range and to interpret them on the basis of coventional dielectric theory.

§ 2. EXPERIMENTAL METHOD

Measurements were made of the real (ϵ') and imaginary (ϵ'') parts of the dielectric constant at wavelengths of 3·23 cm, 9·90 cm, 17·2 cm and 51·5 cm. The experimental techniques have been fully discussed previously in connection with other investigations, so that only a brief description is necessary.

At 3·23 cm both alcohols exhibit high absorption, and reflected waves are damped out by a short length of liquid column, thereby leaving a travelling wave in the major part of the specimen. Thus by making measurements of the change in phase and amplitude through the alcohol it was possible to determine its propagation constant for this radiation. The waveguide apparatus together with details of the technique has been described by Buchanan (1952). A coaxial line apparatus based on the same principle was employed at 9·90 cm and 17·2 cm. At 51·5 cm, where the attenuation coefficients of the alcohols are lower, measurements were made on the standing wave set up in a short-circuited line which led to the calculation of the propagation constant (Buchanan and Grant 1955, Grant 1955). A schematic diagram illustrating the methods is given in figure 1 and the results are shown in tables 1 and 2.

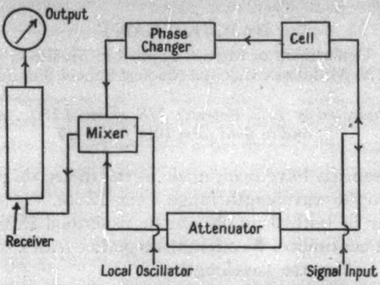


Figure 1. Schematic diagram of the apparatus.

Table 1. Dielectric Properties of Methyl Alcohol at 20°c

Present Measurements			Other Workers		
λ(cm)	ε′ ± 2%	ε″ ± 2%	λ(cm)	ε′	ε″
3·23	8·0	8·15	0·62†	6·04	3·15
9·9	18·7	14·0	0·80‡	5·68	3·23
17·2	25·8	11·6‖	1·24†	6·88	4·76
51·5	31·8	5·39	1·25‡	5·98	4·48
			3·20‡	7·78	7·69
			3·21†	8·32	8·15

Table 2. Dielectric Properties of Ethyl Alcohol at 21°c

Present Measurements			Other Workers		
λ(cm)	ε′ ± 2%	ε″ ± 3%	λ(cm)	ε′	ε″
3·23	4·49	2·37	0·62†	3·47	1·11
9·9	6·0‖	6·1	0·80‡	3·89	1·30
17·2	9·0	9·2	1·24†	4·23	1·56
51·5	18·6	9·0	1·25‡	4·13	1·42
			3·20‡	4·59	2·50
			3·21†	4·54	2·28
			12·6§	5·5	7·8

(The temperatures of both investigations were chosen in order to give a direct comparison with the results of Poley (1955).)

† Lane and Saxton 1952 (results at 21°c obtained by interpolation).

‡ Poley 1955.

§ Bolton 1948.

‖ ± 3%.

Since both alcohols very readily absorb moisture from the atmosphere, the elimination of water presents a serious practical problem, particularly as the dielectric constant of water is much greater than that of either of the alcohols. Consequently, the following precautions were taken. Freshly distilled alcohol which was stated to contain less than 0·1% water was used, and measurements made at 3·23 cm (the shortest wavelength being the most critical for detecting a hydrated sample). Continuous readings were taken until the consistency deteriorated, showing that water was being absorbed. For the particular atmosphere in which the experiment was being conducted, the time before inconsistency started to set in was considerably greater than the time taken for a normal reading, using any of the techniques described. A fresh sample was then taken and ϵ' and ϵ'' determined at 10 cm, 17 cm and 50 cm where the effect of moisture is much less serious.

The above procedure was carried out for both alcohols and it is felt that, in all cases, the moisture content was less than 0·5% during the course of the measurements. The overall limits of error in ϵ' and ϵ'' are shown in tables 1 and 2.

§ 3. ANALYSIS OF RESULTS

3.1. *Results Obtained from the Present Work*

For each alcohol the results can be represented by a Cole–Cole circle obtained by plotting ϵ'' against ϵ'. These plots are shown in figures 2 and 3, representing the results for methyl and ethyl alcohol respectively.

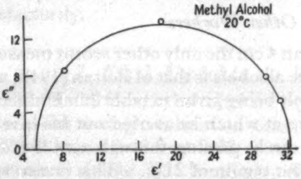


Figure 2. Cole–Cole circle for methyl alcohol at 20°c.

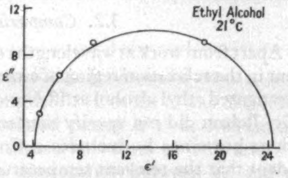


Figure 3. Cole–Cole circle for ethyl alcohol at 21°c.

In each case the fit to the circle is satisfactory with the centre near the axis· thereby showing to a first approximation (Cole and Cole 1941) that the dispersion may be represented by the equation

$$\epsilon = \epsilon' - i\epsilon'' = \epsilon_\infty + \frac{\epsilon_s - \epsilon_\infty}{1 + (i\lambda_s/\lambda)^{1-\alpha}} \qquad \ldots\ldots (1)$$

where ϵ is the complex dielectric constant, ϵ_∞ and ϵ_s are the dielectric constants at frequencies just higher and just lower than the limits of the dispersion region respectively. λ_s is the relaxation wavelength and λ is the free space wavelength of measurement. The parameter α is a measure of the distribution of relaxation times (if existent) and is zero for a single relaxation time.

The applicability of equation (1) can then be checked analytically as has been described previously (Grant, Buchanan and Cook 1947) and it is concluded that α lies in the range $0 \leqslant \alpha \leqslant 0{\cdot}015$ for both alcohols. This shows that the dispersions may be characterized by a single relaxtion time or a slight distribution of relaxation times. As has been shown previously these two possibilities can be distinguished best by measuring ϵ'' at wavelengths well removed from the centre of the dispersion region. Measurements at high frequencies in this region have already been made by Lane and Saxton and by Poley and do, in fact, yield values of ϵ'' higher than would be expected in the case of a single relaxation time, but this is capable of more than one interpretation, as will be discussed later. On the low frequency side of the dispersion regions no recent data on the magnitude of ϵ'' appear to have been published and, as shown by Saxton (1952), there is lack of agreement between the results of the earlier workers.

Consequently, the dispersion in both alcohols can be taken as being characterized either by a single relaxation time or a small distribution of relaxation times, either possibility being equally likely in deductions from the present work.

The dispersion parameters listed in table 3 are obtained on the assumption of a single relaxation time. For reasons to be discussed now, ϵ_∞ is difficult to deduce accurately and, accordingly, limits of error have not been quoted.

Table 3. Dispersion Parameters of Methyl and Ethyl Alcohols

Parameter	$\epsilon_s \pm 1$	λ_s(cm)	ϵ_∞†
Methyl alcohol			
$T=21^\circ$c	32·6(32·2)	10·0 ± 0·3	~5·2
Ethyl alcohol			
$T=21^\circ$c	24·6(24·1)	31 ± 1	~4·2

† Difficult to estimate owing to possible presence of other subsidiary factors. Values in brackets are those of Mecke and Reuter (1949).

3.2. *Comparison with Other Workers*

Apart from work at wavelengths of less than 4 cm the only other recent measurement in the relaxation region of either of these alcohols is that of Bolton (1948) who investigated ethyl alcohol at 12·6 cm, the result being given in table 2. Unfortunately, Bolton did not specify the temperature at which he carried out his investigations but since he includes a Cole–Cole circle passing through $\epsilon_s \simeq 25{\cdot}0$ it is evident that the relevant temperature is in the region of 21°c. This is also supported by Bolton's calculated value of λ_s as 31 cm, in good agreement with the author's value. The other conclusion is that ϵ_∞ takes a value of about 2·5 which is considerably lower than the 4·2 derived from the present work. Saxton (1952) and Poley (1955) also suggest ϵ_∞ for ethyl alcohol as being in the region of 4·0–4·4. The results of Lane and Saxton and of Poley are also included in tables 1 and 2 together with those of Bolton and the author. The results of these other investigators have not been plotted on the circle diagrams (figures 2 and 3) since these figures are not sensitive in revealing small percentage differences at the high frequency end of the dispersion regions, where the majority of the points lie. In addition, the points are not fixed uniquely as a function of wavelength so that a point lying on the circle is not necessarily a better indication of the validity of any particular dispersion theory than one lying just off the circle.

As seen from tables 1 and 2 the agreement between the various workers is not good but nevertheless one feature seems to be common to all the measurements at wavelengths of 3 cm and shorter, namely, the high values of both ϵ' and ϵ'' as compared with those which would be expected if the relaxation were characterized by a single relaxation time. In order to explain these high values, both Saxton and Poley suggest the possibility of the existence of a subsidiary dispersion region and Saxton proposes that this might be a case of resonance absorption as described by Fröhlich (1949). These questions will now be discussed by taking into account the extra evidence provided by the recent measurements at longer wavelengths.

§ 4. THE POSSIBILITY OF DISPERSION AT MILLIMETRE WAVELENGTHS

It will be convenient to consider each alcohol in turn.

4.1. *Methyl Alcohol in the Wavelength Range* 0·62 *cm*–3·3 *cm*

Since ϵ_∞ is not known accurately, and since $\epsilon' \simeq \epsilon_\infty$ in this range, it is clearly not profitable to compare calculated and experimental values of ϵ'. In the case of ϵ'', some estimate of the departure from the Debye equations can be obtained. Thus at 0·62 cm $\epsilon''_{\text{calc}} = 1{\cdot}70$ and $\epsilon''_{\text{exp}} = 3{\cdot}14$ and at 0·802 cm $\epsilon''_{\text{calc}} = 2{\cdot}26$ and $\epsilon''_{\text{exp}} = 3{\cdot}23$ (table 1). Similar, but smaller, divergencies occur at 1·25 cm, while at 3·2 cm the calculated and measured values agree to within experimental error. As stated by Saxton, a very wide distribution of relaxation times could be invoked to explain these high readings, but this would be unlikely. The present work supports this unliklihood since, if a distribution of relaxation times does exist in the main dispersion region then it is certainly very narrow.

Now since the infra-red value of ϵ' has been found to be 1·98 at 152 μ (Cartwright and Errera 1936) and since ϵ_∞ has been found by most workers to lie between 5 and 6, there must be further dispersion in a region in the wavelength range 152 μ–0·62 cm. Saxton proposed that this further dispersion be centred around a wavelength of about 2·5 mm. This would then account for his experimental facts satisfactorily.

4.2. *Ethyl Alcohol in the Wavelength Range* 0·62 *cm*–3·3 *cm*

The results for ethyl alcohol are equally inexplicible in terms of one dispersion region with a single relaxation time, the calculated values of ϵ'' at 0·62 cm and 0·802 cm being 0·40 and 0·52 as against the measured values of 1·11 and 1·30 respectively (table 2). Similarly, there is a discrepancy between ϵ_∞ and the magnitude of ϵ' at 152 μ, the former lying between 4·0 and 4·4 and the latter taking a value of about 2·16. In the same way as for methyl alcohol, Saxton suggests the possibility of resonance absorption, with the peak in this case occurring at a wavelength of about 5 mm.

Unfortunately, there appear to be significant divergencies between the results of Lane and Saxton and those of Poley as shown in tables 1 and 2. For example, in the case of methyl alcohol, Lane and Saxton record $\epsilon' = 6{\cdot}04$ at 0·62 cm and Poley's value of ϵ' is 7% *less* than that of Lane and Saxton although his free-space wavelength is *greater* by about 30%. Also, if a direct comparison is made between the results obtained at 1·24 cm (Lane and Saxton) and those at 1·25 cm (Poley) it is seen that the two values of ϵ' differ by 15%. Since the authors in each case claim their results accurate to about ±2%, it seems that a systematic error has arisen.

In the case of ethyl alcohol, the agreement is better than for methyl alcohol except for Bolton's measurement which yields a value of ϵ_∞ equal to about 2·5. Although Bolton claims ±1% in both ϵ' and ϵ'' it rather appears that an error has arisen since the results of the other workers (including the author) yield values of ϵ_∞ much higher than this.

Thus it is clear from the above that too much reliance should not be placed on the exact magnitudes of ϵ' and ϵ'' of either alcohol in this wavelength region of 0·62 cm–3·3 cm. It is equally clear, however, that even when taking into account all possible errors, the experimentally determined values of ϵ' and ϵ'' are still too high to be explained solely in terms of one dispersion region with a single relaxation time and, although a small distribution of relaxation times is possible, this would not account satisfactorily for the observations. Thus it may be concluded that Saxton's suggestion regarding a further dispersion region is correct qualitatively, although not necessarily quantitatively.

4.3. *Further Suggestions Regarding the Nature of the Subsidiary Absorption Region*

For both alcohols, measurements in the infra-red in the wavelength range 52 μ–152 μ have been reported (Cartwright and Errera 1936) and it is of interest to attempt to link up these results with the present work. (No investigations appear to have been made at any wavelength between 152 μ and 0·62 cm.) Cartwright and Errera's results are shown in table 4.

Table 4. Behaviour of Methyl and Ethyl Alcohols in the Infra-red †

Free-space wavelength (μ)	52	63	83	100	117	152
Methyl Alcohol						
Refractive index n	1·26	1·28	1·34	1·39	1·41	1·41
Absorption coefficient k	0·085	0·117	0·173	0·171	0·187	0·235
Ethyl Alcohol						
Refractive index w	1·23	1·34	1·37	1·42	1·47	1·48
Absorption coefficient k	0·041	0·052	0·085	0·088	0·089	0·108

† Measurements of Cartwright and Errera (1936)

It is evident from table 4 that in the infra-red, both ϵ' and ϵ'' are increasing towards longer wavelengths, thereby suggesting the beginning of a dispersion region. Also, the high values of ϵ'' recorded at wavelengths in the range 0·62 cm–1 cm suggest that a dispersion region is also beginning in *this* range and extending towards higher frequencies.

What is now proposed is that these two above effects are, in fact, the opposite ends of the same dispersion region. This is quite possible since the frequency ratio between 152 μ and 0·62 cm is only about 40 or 1·0 decades which would be acceptable for a dispersion of this type. Saxton suggested that the absorption was centred around 2·5 mm for methyl alcohol and 5 mm for ethyl alcohol, but if this were so then the contribution to ϵ'' in the infra-red would be negligible which in practice does not appear to be the case, since $\epsilon'' = 2nk$ (table 4). The other alternative (assuming Saxton to be correct) would be to postulate the existence of *two* absorption regions, one in the infra-red and the other in the millimetre range. This is possible, but owing to the smaller frequency range which each dispersion region would occupy, is not as likely as one region with absorption maxima centred around, say, 1 mm for methyl alcohol and 2 mm for ethyl alcohol. If this latter possibility were correct, then the contribution to ϵ'' in the wavelength range 0·62 cm–1 cm would probably not account for the whole of the recorded discrepancy between the measured values of ϵ'' and those calculated on the basis of only one main dispersion region characterized by a single relaxation time.

This is not a serious point, however, since, as pointed out in the previous section, appreciable experimental errors may be involved in the recorded values of ϵ''.

An alternative method of presenting the results is shown in figure 4 which illustrates for both alcohols a plot of tan δ against the logarithm of wavelength.

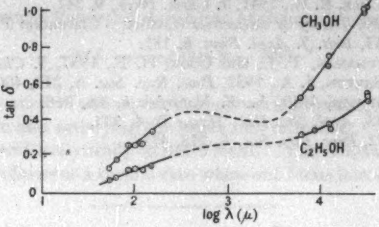


Figure 4. Plot of tan δ against log λ. ($\tan\delta = \epsilon''/\epsilon'$).

From the inflection in the curve for methyl alcohol it is evident that dispersion is occurring in the neighbourhood of 1 millimetre. For ethyl alcohol the information furnished by the graph is less conclusive.

Precise evidence in favour of any one of the above possibilities cannot be advanced until accurate readings are available at wavelengths in the neighbourhood of 2 mm.

§ 5. CONCLUSIONS

The principal conclusions can be listed as follows.

(*a*) Both methyl and ethyl alcohols exhibit dipolar relaxation in the decimetre wavelength range, with relaxation wavelengths at about 20°C equal to 10 cm and 31 cm respectively. The dispersions may be characterized by a single relaxation time or a small distribution of relaxation times.

(*b*) For both alcohols in the wavelength range 0·62 cm–3 cm the experimental results obtained by other investigators (Lane and Saxton 1952, Poley 1955) do not agree with the theoretical values calculated from the data obtained in the main part of the relaxation regions and therefore it appears that another subsidiary dispersion mechanism is coming into play. It is suggested for both alcohols that this subsidiary dispersion region constitutes the low frequency end of the region appearing at a wavelength of about 100 microns and experimentally observed by Cartwright and Errera (1936).

§ 6. ACKNOWLEDGMENTS

The author is indebted to Dr. H. F. Cook for his guidance and interest in this work, and to Dr. J. Ph. Poley for suggesting the investigations. Thanks are also due to a referee of the paper for advising the inclusion of the plot illustrated in figure 4.

This work formed part of a research programme leading to the Ph.D. degree of the University of London. It gives pleasure to thank the Medical Research Council for a grant during the course of the investigations.

REFERENCES

BOLTON, H. C., 1948, *Proc. Phys. Soc.* B, **91,** 294.

BUCHANAN, T. J., 1952, *Proc. Instn Elect. Engrs*, **99,** 61.

BUCHANAN, T. J., and GRANT, E. H., 1955, *Brit. J. Appl. Phys.*, **6,** 64.

CARTWRIGHT, C. H., and ERRERA, J. 1936, *Proc. Roy. Soc.* A, **154,** 138.

COLE, K. S., and COLE, R. H., 1941, *J. Chem. Phys.*, **9,** 341.

FRÖHLICH, H., 1949, *Theory of Dielectrics* (Oxford : Clarendon Press).

GRANT, E. H., 1955, *Brit. J. Appl. Phys.* **6,** 181.

GRANT, E. H., BUCHANAN, T. J., and COOK, H. F., 1957, *J. Chem. Phys.*, **26,** 156.

LANE, J. A., and SAXTON, J. A., 1952, *Proc. Roy. Soc.* A, **213,** 400.

MECKE, R., and REUTER, 1949, A., *Z. Naturforsch*, **4A,** 368.

POLEY, J. PH., 1955, *Appl. Sci. Res., Hague* B, **4,** 337.

SAXTON, J. A., 1952, *Proc. Roy. Soc.* A, **213,** 473.

The Two-Mirror Aplanat

By A. K. HEAD

Aeronautical Research Laboratories, Melbourne, Australia†

MS. received 18th June 1957

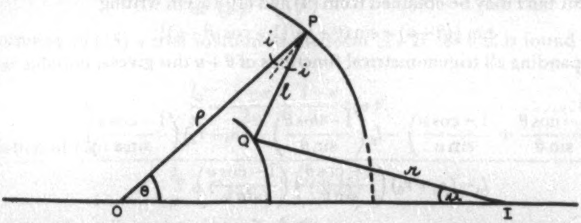
Abstract. A combination of two aspherical mirrors is considered which is aplanatic (i.e. axially stigmatic and satisfying the exact sine criterion). The shapes of the mirrors are determined analytically in finite form. The Schwarzschild–Chretien aplanatic telescope follows as a special case when one focus is at infinity.

§ 1. INTRODUCTION

SCHWARZSCHILD (1905)‡ and Chretien (1922) have considered the combination of two aspherical mirrors to form an aplanatic (i.e. spherically correct and coma free) telescope and have given an analytical solution in finite form for the shapes of the mirrors. Bureau and Swings (1934) have attempted to solve the extended problem of an aplanatic mirror-pair with foci at finite distances. Their solution is in the form of infinite series and is apparently of limited use as Burch (1947) has used the Schwarzschild–Chretien telescope in reverse as the basis of the design of his reflecting microscope objective and has experimentally corrected this to work at finite tube length. An analytical solution in finite form is given here for the two-mirror aplanat working between finite foci.

§ 2. ANALYSIS

The figure shows an axial section of the system and the path OPQI of a typical ray from object O to image I. The first mirror is specified by polar coordinates (ρ, θ) about O and the second mirror by polar coordinates (r, u) about I. It is convenient to refer to these as the ρ-mirror and r-mirror respectively. The angle of incidence and reflection of the ray at the ρ-mirror is i and the distance PQ is l. The poles of the mirrors are at $\rho = \rho_0$ and $r = r_0$ and the distance between the poles of the mirrors is l_0. Thus the distance between O and I is $\rho_0 + r_0 - l_0$.



---

† Now at the Division of Tribophysics, Commonwealth Scientific and Industrial Research Organization, University of Melbourne, Australia.

‡ This reference was not available to the author and is quoted from Burch (1947).

The system is then specified by the following equations (1) to (5).

Axial stigmatism:

$$\rho + r + l = \rho_0 + r_0 + l_0 \qquad \ldots\ldots(1)$$

Sine condition:

$$\sin\theta = m\sin u \qquad \ldots\ldots(2)$$

Reflection at ρ-mirror:

$$\frac{1}{\rho}\frac{d\rho}{d\theta} = -\tan i \qquad \ldots\ldots(3)$$

Projection of PQ parallel and perpendicular to axis:

$$l\cos(2i+\theta) = \rho\cos\theta + r\cos u - (\rho_0 + r_0 - l_0) \qquad \ldots\ldots(4)$$

$$l\sin(2i+\theta) = \rho\sin\theta - r\sin u. \qquad \ldots\ldots(5)$$

From equations (1) (2) (4) and (5) the variables r, l and u will be eliminated and the resulting expression for $\tan i$ substituted in (3) to give a differential equation connecting ρ and θ.

Multiplying (5) by $\cos u$ and (4) by $\sin u$ and adding gives

$$l\sin(2i+\theta+u) = \rho\sin(\theta+u) - (\rho_0 + r_0 - l_0)\sin u \qquad \ldots\ldots(6)$$

and multiplying (4) by $\cos u$ and (5) by $\sin u$ and subtracting gives

$$l\cos(2i+\theta+u) = \rho\cos(\theta+u) + r - (\rho_0 + r_0 - l_0)\cos u, \qquad \ldots\ldots(7)$$

i.e.

$$l + l\cos(2i+\theta+u) = \rho\cos(\theta+u) - (\rho_0 + r_0 - l_0)\cos u - \rho + (\rho_0 + r_0 + l_0) \qquad \ldots\ldots(8)$$

on substituting for r from (1).

Division of (6) by (8) gives

$$\frac{\sin(2i+\theta+u)}{1+\cos(2i+\theta+u)} = \frac{\rho\sin(\theta+u) - (\rho_0 + r_0 - l_0)\sin u}{\rho\cos(\theta+u) - (\rho_0 + r_0 - l_0)\cos u - \rho + (\rho_0 + r_0 + l_0)}. \qquad \ldots\ldots(9)$$

Now

$$\frac{\sin(2i+\theta+u)}{1+\cos(2i+\theta+u)} = \tan\{i + \tfrac{1}{2}(\theta+u)\} = \frac{\tan i + \tan\frac{1}{2}(\theta+u)}{1 - \tan i\tan\frac{1}{2}(\theta+u)}. \qquad \ldots\ldots(10)$$

Thus $\tan i$ may be obtained from (9) and (10). On writing

$$\tan\tfrac{1}{2}(\theta+u) = \sin(\theta+u)/\{1+\cos(\theta+u)\}$$

and expanding all trigonometrical functions of $\theta + u$ this gives

$\tan i =$

$$\frac{\rho\left(\dfrac{1-\cos\theta}{\sin\theta} + \dfrac{1-\cos u}{\sin u}\right) - l_0\left(\dfrac{1-\cos\theta}{\sin\theta}\right) - (\rho_0 + r_0)\left(\dfrac{1-\cos u}{\sin u}\right)}{l_0 - (\rho_0 + r_0)\left(\dfrac{1-\cos\theta}{\sin\theta}\right)\left(\dfrac{1-\cos u}{\sin u}\right)}. \qquad \ldots\ldots(11)$$

From (2) $$\sin u = \frac{\sin\theta}{m}$$

and so $$\cos u = \pm\sqrt{\left(1 - \frac{\sin^2\theta}{m^2}\right)} \qquad \ldots\ldots(12)$$

and $$\frac{1-\cos u}{\sin u} = \frac{m \mp \sqrt{(m^2 - \sin^2\theta)}}{\sin\theta}. \qquad \ldots\ldots(13)$$

To resolve this ambiguity of sign the following convention will be adopted. In (12) the positive sign will be taken so that as $\theta \to 0$, $u \to 0$. This would give

$$\frac{(1-\cos u)}{\sin u} = \frac{m-\surd(m^2-\sin^2\theta)}{\sin\theta} \qquad \ldots\ldots(13\,a)$$

except that m has been taken inside the square root. However (13 *a*) is correct if $\surd(m^2-\sin^2\theta)$ is given the sign of m, i.e.

$$\surd(m^2-\sin^2\theta) = m\left|\sqrt{\left(1-\frac{\sin^2\theta}{m^2}\right)}\right|$$

and this convention will be used.

Substituting (13 *a*) in (11), writing $\kappa = (\rho_0 + r_0)/l_0$ and using (3) gives

$$\frac{1}{\rho}\frac{d\rho}{d\theta} = \frac{(\rho/l_0)\sin\theta\{1-\cos\theta+m-\surd(m^2-\sin^2\theta)\} - \sin\theta\{1-\cos\theta+\kappa m-\kappa\surd(m^2-\sin^2\theta)\}}{\kappa(1-\cos\theta)\{m-\surd(m^2-\sin^2\theta)\}-\sin^2\theta}. \qquad \ldots\ldots(14)$$

On multiplying through by l_0/ρ, (14) can be written

$$\frac{d}{d\theta}\left(\frac{l_0}{\rho}\right) = \left(\frac{l_0}{\rho}\right)F(\theta) + G(\theta) \qquad \ldots\ldots(15)$$

where

$$F(\theta) = \frac{\sin\theta\{1-\cos\theta+\kappa m-\kappa\surd(m^2-\sin^2\theta)\}}{\sin^2\theta-\kappa(1-\cos\theta)\{m-\surd(m^2-\sin^2\theta)\}},$$

$$G(\theta) = -\frac{\sin\theta\{1-\cos\theta+m-\surd(m^2-\sin^2\theta)\}}{\sin^2\theta-\kappa(1-\cos\theta)\{m-\surd(m^2-\sin^2\theta)\}}.$$

This is a first order linear differential equation in l_0/ρ and its solution is the sum of (*a*) any particular solution of (15) and (*b*) an arbitrary multiple of the solution of the corresponding homogeneous equation, i.e. of

$$\frac{d}{d\theta}\left(\frac{l_0}{\rho}\right) = \left(\frac{l_0}{\rho}\right)F(\theta). \qquad \ldots\ldots(16)$$

By substituting in (15) a trial solution of the form $A + B\cos\theta$ it is found that a particular solution is

$$\frac{l_0}{\rho} = \frac{1+\kappa}{2\kappa} + \frac{1-\kappa}{2\kappa}\cos\theta. \qquad \ldots\ldots(17)$$

The solution of (16) is

$$\frac{l_0}{\rho} = C\exp\left\{\int F(\theta)d\theta\right\}$$

and this integral can be evaluated by the substitution

$$\gamma = \cos\theta + \surd(m^2-\sin^2\theta), \qquad \ldots\ldots(18)$$

i.e.,

$$\cos\theta = (\gamma^2-m^2+1)/2\gamma$$

to give

$$\frac{l_0}{\rho} = C[\gamma]^{-1}[\gamma-(1-m)]^{\alpha}[\gamma-(m-1)]^{\beta}[(\kappa+1)\gamma-(\kappa-1)(m+1)]^{2-\alpha-\beta} \qquad \ldots\ldots(19)$$

where for ease of presentation $m\kappa/(m\kappa-1)$ has been replaced by α and $m/(m-\kappa)$ by β. The general solution is the sum of (17) and (19) and the arbitrary constant C is determined by $\rho=\rho_0$ when $\theta=0$. This gives as the equation for the ρ-mirror

$$\frac{l_0}{\rho}=\frac{1+\kappa}{2\kappa}+\frac{1-\kappa}{2\kappa}\cos\theta+$$
$$\left[\frac{l_0}{\rho_0}-\frac{1}{\kappa}\right]\left[\frac{\gamma}{1+m}\right]^{-1}\left[\frac{\gamma-(1-m)}{2m}\right]^{\alpha}\left[\frac{\gamma-(m-1)}{2}\right]^{\beta}\left[\frac{\kappa+1}{2m+2}\gamma-\frac{\kappa-1}{2}\right]^{2-\alpha-\beta} \qquad \ldots\ldots(20)$$

where γ is a function of θ given by (18).

It will be seen from the reversibility of the system that the r-mirror will have a similar equation with m replaced by $1/m$.

Specifically if $M=1/m$ then

$$\frac{l_0}{r}=\frac{1+\kappa}{2\kappa}+\frac{1-\kappa}{2\kappa}\cos u+$$
$$\left[\frac{l_0}{r_0}-\frac{1}{\kappa}\right]\left[\frac{\delta}{1+M}\right]^{-1}\left[\frac{\delta-(1-M)}{2M}\right]^{\alpha'}\left[\frac{\delta-(M-1)}{2}\right]^{\beta'}\left[\frac{\kappa+1}{2M+2}\delta-\frac{\kappa-1}{2}\right]^{2-\alpha'-\beta'} \qquad \ldots\ldots(21)$$

where $\alpha'=M\kappa/(M\kappa-1)$, $\beta'=M/(M-\kappa)$ and $\delta=\cos u+\sqrt{(M^2-\sin^2 u)}=M\gamma$ from (2) and (18).

§ 3. SPECIAL CASES

(*a*) If the object or image is at infinity, (20) and (21) reduce to the Schwarzschild–Chretien formulae. Considering the case where the image is at infinity it is necessary to have

$$r\to\infty,\ \kappa\to\infty,\ m\to\infty,\ m/\kappa\to l_0$$

and the Schwarzschild–Chretien expressions follow.

(*b*) If $m=1$ then

$$\frac{\rho_0}{\rho}=\frac{\rho_0-r_0+l_0}{2l_0}+\frac{r_0-\rho_0+l_0}{2l_0}\cos\theta$$

and

$$\frac{r_0}{r}=\frac{r_0-\rho_0+l_0}{2l_0}+\frac{\rho_0-r_0+l_0}{2l_0}\cos u.$$

These are polar equations of conics about a focus. In general one mirror is an ellipse and the other a hyperbola, but as special cases, both can be parabolas or one a sphere and the other plane.

(*c*) If $m=-1$ then

$$\frac{\rho_0}{\rho}=\frac{\rho_0+r_0+l_0}{2l_0}-\frac{\rho_0+r_0-l_0}{2l_0}\cos\theta$$

and

$$\frac{r_0}{r}=\frac{\rho_0+r_0+l_0}{2l_0}-\frac{\rho_0+r_0-l_0}{2l_0}\cos u.$$

As in (*b*) these are polar equations of conics about a focus, but in this case both mirrors are geometrically similar. Thus for a system with unit magnification ($m=\pm 1$) the mirrors must be conics.

(*d*) For paraxial rays the radius of curvature of the ρ-mirror is

$$2ml_0\rho_0/(m\rho_0+ml_0-r_0)$$

and of the r-mirror is

$$2l_0r_0/(r_0+l_0-m\rho_0).$$

ACKNOWLEDGMENTS

This paper is published with the permission of the Chief Scientist, Department of Supply.

REFERENCES

BURCH, C. R., 1947, *Proc. Phys. Soc.*, **59**, 41.
BUREAU, F., and SWINGS, P., 1934, *Rev. Opt.* (*Théor. Instrum.*), **13**, 127.
CHRETIEN, H., 1922, *Rev. Opt.* (*Théor. Instrum.*), **1**, 49.
SCHWARZSCHILD, 1905, *Theorie der Spiegel Teleskop* (Göttingen Observatory).

Stress Annealing in Vacuum Deposited Copper Films

By H. S. Story† and R. W. Hoffman
Case Institute of Technology, Cleveland, Ohio

MS. received 13*th August* 1956, *and in revised form* 19*th June* 1957

Abstract. An apparatus has been built for the measurement of mechanical stress in vacuum deposited films. A study of the stress changes during pulse annealing has been made with copper films deposited at several substrate temperatures between $-150°$c and $+75°$c. The annealing spectrum shows a prominent peak centred near room temperature which matches a resistance annealing peak in similar films and also in cold worked copper. The fall of tensile stress in this region suggests a vacancy mechanism. The theory of Murbach and Wilman for the behaviour of initial stress with temperature is not supported by these results.

§ 1. Introduction

Previous work has shown that vacuum deposited films are in a state of high mechanical stress (Hoffman, Anders and Crittenden 1953, Hoffman, Daniels and Crittenden 1954, Murbach and Wilman 1953, Wilman1955, Hoffman and Story 1955).

When a film is deposited on a thin glass or mica substrate, the stress manifests itself in the curvature of the film and substrate system. The magnitude of the stress depends on the substrate temperature during film formation and on subsequent annealing treatment. Relative to its variation with temperature, the stress may be thought of as consisting of two parts : reversible and irreversible. The reversible part comes about as a result of differential expansion between film and substrate. It is defined as zero when the film and substrate are at the deposition temperature. The irreversible part, referred to as intrinsic stress, is presumed in this paper to result from the presence of lattice defects in the film. Irreversible change in stress takes place upon annealing at temperatures higher than the deposition temperature, and it is with this effect that we are principally concerned.

In previous experiments in this laboratory the films were made in one vacuum system and then transferred to another for the stress measurements. For this reason, the work was limited to the range above room temperature.

The purpose of this paper is to describe an apparatus which extends the stress measurements to the range below room temperature and to present the results of such measurements on copper films in the deposition temperature range $-150°$ to $+75°$c. Stress measurements are carried out in the same system in which the film is made and thus the film is never exposed to the atmosphere.

A related quantity of interest is the electrical resistivity of vacuum deposited films. The resistivity of a film is usually higher than that of the bulk material for two reasons (Sondheimer 1952, Crittenden and Hoffman 1953, Rasor

---

† Now at Bell Telephone Laboratories, Whippany, N.J.

1955): first, the surfaces of the film impose a limitation on the free paths of conduction electrons (this effect will not be considered further here); secondly, the presence of lattice defects in the film also limits the free paths; this contribution to the resistivity can undergo irreversible change upon annealing. By comparison between stress and resistivity annealing, insight into annealing mechanisms can be gained. The annealing of these properties is related to the annealing of resistivity following cold work, and the important question of the volume change can be answered for at least one of the annealing regions which are found.

§ 2. EXPERIMENTAL TECHNIQUE

2.1. *General*

The experiment consists of depositing a film on a thin mica substrate which is clamped in a system of copper blocks. Clamping facilitates heat transfer from the substrate. After the film is made, the substrate is partially released, remaining clamped only at one end. The film and substrate system may now be thought of as a cantilever beam which tends to curve like a bimetallic strip. This tendency to curve is counteracted by a force applied to the free end of the beam by an electrodynamic balance. The stress measurement consists of determining the applied force as a function of the position of the end of the beam.

So that the rate of deposition and thickness of the sample film will be known, the electrical conductance of another film on a monitor substrate is recorded.

2.2. *Vacuum System and Thickness Measurements*

The measurements are made on films formed in a metal vacuum system by techniques similar to those described in detail elsewhere (Crittenden and Hoffman 1953, Hoffman, Daniels and Crittenden 1954). The system consists of a steel bell jar and baseplate with 27 rubber gasketed electrical feed-throughs and tubing fittings. Pump out is accomplished with a Kinney roughing pump and a D.P.I. MCF300 diffusion pump backed by a Megavac forepump.

The geometry of the parts in the vacuum system is shown schematically in figure 1. The metal to be studied is electrodeposited on a tungsten filament which had been previously outgassed. On heating *in vacuo*, the filament forms a source of metal vapor which condenses on to two substrates located at equal distances from the source. One substrate is the stress sample and will be described presently. The other substrate is a monitor which provides a measure of the thickness of the stress sample film. It is a section cut from a glass microscope slide with painted and fired silver and platinum probes to which the monitor film makes electrical contact. The conductance of the monitor film is continuously recorded as the deposition proceeds. Mechanical shutters between the filament and the sample substrate start and stop the deposition of the sample film. Times of opening and closing are recorded on the monitor film conductance record. The thickness of the sample film is proportional to the change of conductance which occurs between the opening and closing of the shutters.

The monitor substrate is maintained at a temperature high enough to give rapid resistance annealing in the metal being studied. In the case of copper, the temperature chosen is +100°c. Since the conductivity of the material as it is depositing is not well known, there is need for an independent calibration. This is provided by a chemical determination of the mass of material deposited on the sample substrate. The film is dissolved off the substrate and the metal is converted to a coloured ion. For the copper work, the sodium diethyldithiocarbamate reaction (Sandell 1944) was used and the optical density determined with a Beckman Model B spectrophotometer.

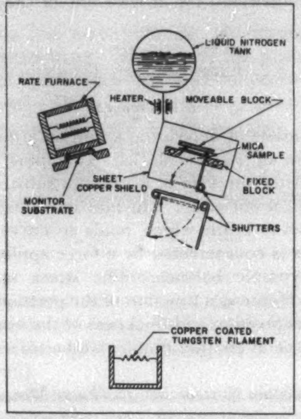


Figure 1. The geometry of the parts inside the vacuum system.

Before use, the monitor substrate is put through a standard washing procedure. This consists of a mechanical cleaning with cotton in a boiling solution of commercial detergent (Nacconol NR) and several hot distilled water rinses. The substrate is then dried using commercial paper towels or Cenco bibulous paper.

2.3. *The Stress Apparatus*

The stress sample and measurement apparatus are shown in figures 1 and 2. The sample substrates used in these experiments were mica slabs of thickness 0·003 to 0·005 cm (the variation is from one sample to another), width 0·5 cm and length 7·5 cm cut with the long dimension parallel to the mica crystallographic *b* axis.

In early work, the samples were put through the same cleaning as the monitor strip. It was found however that electrostatic charges (due principally to the drying of the mica) interfered with the stress measurements. To correct this, thin layers of nickel (about 100 Å) were vacuum deposited on to both sides of the mica after the cleaning. The use of thin layers on both sides minimizes the effect of stress in these layers. Before use, in a run, the substrates were annealed overnight at 200°c or higher. Since a number of such samples were made simultaneously, they were stored in a vacuum system or atmospheric pressure desiccator until used.

In figure 2, the important details of the stress apparatus are shown. It consists of three copper blocks, a modified D'Arsonval meter movement and an optical system. Not shown is a copper shield which surrounds the blocks. The right-hand end of the mica substrate is permanently clamped to the fixed block. During deposition, the movable block is lowered on to the fixed block by a clamping mechanism. Along the bottom of the movable block are four milled channels into which lead fuse wires are imbedded. The two inner channels are so spaced that when the movable block is lowered, the fuse wires clamp the edges of the substrate to the fixed block. When the block is in this position, there is sufficient heat transfer between the substrate and blocks so that the temperature rise of the substrate does not exceed 5°c during deposition. (This was determined by resistance measurements on annealed films. The resistance was monitored while a blank filament was heated to the evaporating temperature.) Deposition takes place on the bottom of the substrate through a slot in the fixed block. The dimensions of the film are defined by the machined edges of the slot.

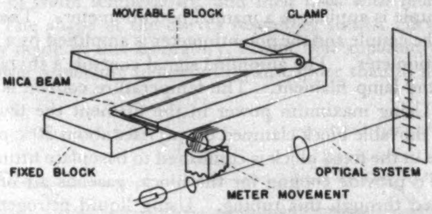


Figure 2. The sample holder and stress measuring apparatus.

For a stress measurement the movable block is raised, freeing the substrate everywhere except at the right-hand end. In nearly all cases the film is found to be in tension, which would tend to curl the substrate downward. This is prevented by the application of a force at the free end of the substrate. The force is applied by a lever attached to the coil of an electromagnetic balance. This balance is a small D'Arsonval meter movement whose restoring torque has been removed by the reversal of one of the springs. The force available is proportional to the current in the meter coil and independent of the angle of the coil. The image of a counter arm also attached to the coil is focused on the scale by the optical system. The position of the free end of the substrate is thus indicated by the scale reading. In each stress measurement, a number of deflections and corresponding currents are recorded. The resulting relation between current and deflection is found and extrapolated to zero deflection. From a similar measurement carried out before the film was deposited, the characteristics of the blank beam are found. The stress is then proportional to the difference in current necessary to give zero deflection with and without the film. From the work of Hoffman (1949) and Brenner and Senderoff (1949), it can be shown that the stress is given by

$$S = 4LJi/3dWtl$$

where L, W and t are the length, width and thickness of the film, J is a meter calibration constant, l is the length of the meter arm, d is the substrate thickness and i is the difference in current. J was determined by loading the meter arm with known weights. W was measured with a travelling microscope and d was measured with a Geneva comparator. The current in the meter coil was determined by means of a Speedomax recorder reading the potential difference across a known resistor in series with the coil. No calibration of the optical system was required since only a null reading was used.

To provide some thermal isolation for the blocks and substrate, they are surrounded by a copper shield. A slot in the bottom of the shield permits passage of the depositing atoms. After deposition a shutter is closed to cover this opening. The temperature of the blocks is measured by a chromel–alumel thermocouple in the fixed block. Similar thermocouples give the temperatures of the shield and the monitor furnace.

Heat is supplied to the shield and the monitor furnace by manually controlled resistance heaters. The blocks are heated by a 100 watt lamp filament in a hole which runs the length of the fixed block. The power dissipated by this heater can be controlled manually or automatically. For automatic control the e.m.f. of the thermocouple is applied to a manual potentiometer. The small difference signal between the couple and the potentiometer is amplified by a General Electric recording potentiometer. The amplified signal controls a thyratron tube whose output powers the lamp filament. The temperature control is normally better than $\pm 0{\cdot}5$°c. Using maximum power in the filament the temperature of the blocks (with the movable block clamped down) rises about 30°c per minute.

Another hole in the fixed block is connected to baseplate fittings with stainless steel tubing. To provide cooling for the block, gaseous air or liquid nitrogen may be circulated through this tubing. Using liquid nitrogen, a cooling rate greater than 30°c per minute is possible above -150°c. Air or liquid nitrogen may also be passed through the shield tank. Since -150°c is the lowest shield temperature that can easily be obtained, the measurements have been restricted to the range above this temperature.

2.4. *Nature of the Films and their Treatment*

The results in this paper are concerned with copper films in the thickness range 900–1500Å. They were formed at a nominal rate of deposition of 60Å per second. The ‘ exact ’ rate varied from 38 to 70Å per second from one film to another, but was constant within a few per cent for any given film. Residual pressure in the vacuum system was 1×10^{-6} mm Hg but the pressure rose during deposition to about 4×10^{-6} mm. The increase is believed to be due to the release of gas from the electroplated metal on the filament. This pressure would lead to a maximum of 2% contamination by foreign gas atoms if all those which strike the film stick to it (Heavens 1955). No correlation between the magnitude of the stress or stress change and the pressure during deposition has been found.

Another type of contamination would occur if appreciable film oxidation took place during annealing. As a check on this some films were covered with a layer of SiO evaporated from a supplementary filament. There is no significant difference between coated and uncoated films as far as stress behaviour is concerned.

After deposition, the film is taken through a tempering treatment in which it is pulsed to successively higher temperatures and returned to a base temperature for the stress measurement. The time on the high temperature is 15 minutes and each anneal is 25°c higher than the previous one. The overall rate is thus 100°c per hour if down time is not counted.

To eliminate differential expansion effects between the film and substrate, the base temperature usually used is the deposition temperature. When the up and down pulses become too long (several hundred degrees), this becomes impracticable and a shift is made to a higher base temperature. When the change is made, the stress is measured at both the old and new base temperatures, and the difference in stress is used to correct all subsequent values of stress back to the deposition temperature.

§ 3. Experimental Results

In figures 3, 4, 5, 6 and 7 the stress remaining at the end of each 15 minute anneal is plotted at the annealing temperature. Separate curves are shown for films made at different deposition temperatures. In general, the stress is tensile and starts very high for deposit temperatures below 0°c, falls to a minimum in the neighbourhood of room temperature, and then rises with increasing annealing temperature. The shape of the curves resembles similar curves for the nickel films previously made in this laboratory, except that apparently related changes take place at lower annealing temperatures in copper than in nickel.

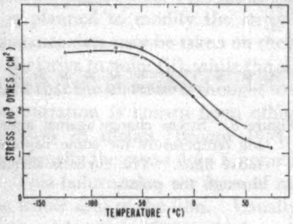


Figure 3. Stress against annealing temperature for a film deposited at -150°c and a film deposited at -125°c.

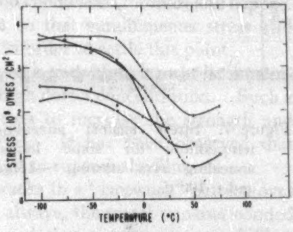


Figure 4. Stress against annealing temperature for films deposited at -100°c, five runs. No reason is known for the variation between runs.

Corresponding resistance–annealing curves for copper are shown in figure 9. These include data taken by N. S. Rasor in this laboratory on a different set of films (Rasor 1955), and some more recent data. Like the stress, the high initial resistivity also anneals out irreversibly as the temperature is raised.

In figures 8 and 10 the quantities ΔS and ΔR are plotted for some representative films. ΔS and ΔR are the change in stress and resistivity respectively which took place at each anneal. These curves are thus proportional to the slopes of the curves of stress and resistivity plotted against annealing temperature.

The resistance and stress annealing picture for copper may be summarized in the following way. There are three regions into which the annealing process is divided : A, a region between -150°c and -25°c in which the stress change is

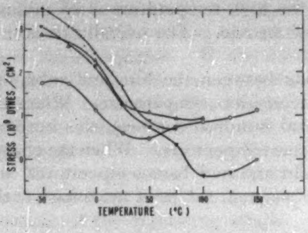


Figure 5. Stress against annealing temperature for films deposited at −50°c, four runs.

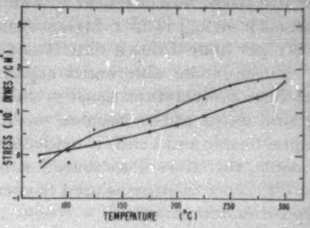


Figure 6. Stress against annealing temperature for films deposited at +75°c. The film marked by crosses was measured by Daniels in the earlier stress apparatus.

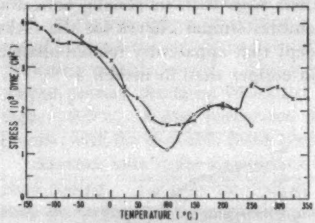


Figure 7. Stress against annealing temperature for some longer annealing runs starting −150°c and −100°c.

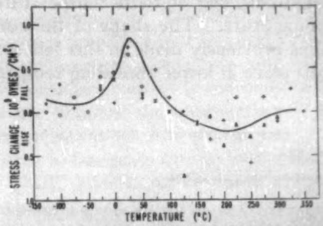


Figure 8. Stress change against annealing temperature for some representative films. The curve is sketched through the points.

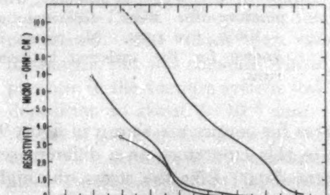


Figure 9. Resistivity against annealing temperature for films made at −150°c and −100°c from the data of Rasor, and a film made at −50°c from more recent data. Sketched from data points at 12·5°c intervals.

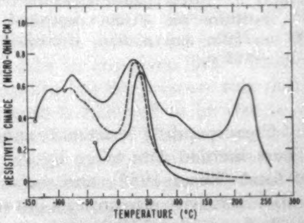


Figure 10. Resistivity change against annealing temperature for the films in figure 9. The −100°c curve is dotted for clarity.

quite small and the resistance change is large ; B, a region between −25°c and +50°c in which both the stress and resistance changes are large ; and C, a region above +50°c in which the stress change is large and the resistance change is small.

There seems to be little doubt that the stress and resistivity changes in region B may be traced to the same mechanism. The peaks in the two slope curves both occur at the same temperature, i.e. about +25°c, and have the same general shape. This region will be considered in more detail below.

The connection between stress and resistivity annealing for regions A and C is not clear. The lack of stress change in region A could be explained in several ways : (i) the temperature of the substrate does not remain near the block and shield temperature during the stress measurement and ‘ pre-annealing ’ takes place ; (ii) the metal cannot support stresses larger than those observed and accommodates itself by yielding ; (iii) an imperfection process operates which has a larger effect on resistivity than on stress. It is not at present possible to decide with certainty among these choices.

In order to investigate point (i), stress measurements were made on well annealed films to determine the differential temperature coefficient of expansion between the metal and the substrate. The stress was measured with the block and shield held at the same temperature. This was repeated for several temperatures over the range being studied. The fact that the stress was linear with temperature indicated that the substrate did remain close to the block in temperature during the measurement time. The measured differential coefficient compares within the experimental error with the known expansion coefficients for mica and copper. It is planned to modify the stress apparatus so that simultaneous stress and resistance data may be taken on the same film in order to settle this point.

Relative to point (ii), while the stress involved is very high, it must be remembered that the material is thought to contain a high defect concentration. Such a concentration is known from other experiments to increase the strength and hardness of metals (Broom 1954, Glen 1955). It has been observed that occasionally the stress does become large enough to rupture the film.

This failure takes the form of many fine cracks in a ‘ lacework ’ pattern over the whole area of the film. Usually, but not always, the metal remains bonded to the substrate. Other workers have observed that vacuum deposited films have a rupture strength greatly exceeding that of the bulk material (Beams, Breazeale and Bart 1955).

Considering point (iii), no simple imperfection annealing process has been found which can account for the changes observed in region A. Not enough is known about the effect of complex imperfection mechanisms on the volume or resistivity of a metal.

The origin of effects noted in region C is also obscure. Dislocation formation from vacancy aggregates has been suggested for a similar region in nickel stress annealing (Hoffman, Anders and Crittenden 1953). It is also possible that the stress rise is related to recrystallization phenomena. The discussion which follows will therefore be confined to region B, after considering some related experimental results.

§ 4. DISCUSSION

A large amount of work has been done by other workers in the related fields of cold work (Broom 1954) and radiation damage (Glen 1955) annealing in metals.

In cold work experiments with copper, several investigators have found two predominant peaks in the annealing curves, one near $-100°$c and one near room temperature. These peaks fall very close to those found in the films. In radiation damage work, these peaks occur at other temperatures, e.g. the upper peak is in the neighbourhood of $-20°$c. There is no general agreement on the nature of the processes involved in either type of annealing. The resistivity will fall if any type of imperfection is annihilated in the sample or migrates to the surface. Among the simpler processes which have been suggested are : vacancy migration, interstitial migration, vacancy–interstitial annihilation and vacancy coalescence. The vacancy mechanisms are favoured in the room temperature range on the basis of the agreement between estimated and measured activation energies for vacancy motion.

Now, considering the films, the change in region B is the sharpest and best defined one in the spectrum. It undoubtedly corresponds to the peak found in cold work experiments at the same temperature. The stress experiments establish this process as one corresponding to a volume increase of the material (as measured by the stress experiment). Of the four processes listed above, two of them lead to such a volume increase : vacancy migration to the surface and vacancy coalescence.

It is instructive to consider the magnitudes of the changes and their possible relationship to the imperfection mechanism. A very simple model is that of a vacancy migrating to the sample surface. This is equivalent to a vacancy being filled by an atom from the surface. It is assumed that the disappearance of the surface atom does not affect the volume as measured by the stress experiment. The following information is pertinent. There have been several estimates of the additional resistivity which results from a dilute concentration of vacancies in copper (Blatt 1955, Van Buren and Jongenburger 1955, Jongenburger 1955, Overhauser and Gorman 1956). The most recent estimate (Overhauser and Gorman 1956) is 1·5 micro-ohm cm per atom per cent (10·5 micro-ohm cm per atom per cent for interstitials).

Tucker and Sampson (1954) and Eshelby (1954) have estimated that the addition of one atom per cent of vacancies would cause a linear strain of about 0·2%, and that the addition of 1% of interstitials would cause a strain of about 1% in copper.

Using the value of Young's modulus for copper of $1{\cdot}2\times10^{12}$ dyn cm^{-2} yields a stress fall of $2{\cdot}4\times10^{9}$ dyn cm^{-2} due to the filling of one per cent vacancies.

In the region B, about three micro-ohm cm of resistivity and 2×10^{9} dyn cm^{-2} of stress are observed to anneal out. Thus, of the order of one or two per cent vacancies have disappeared in this region. Not much more than order of magnitude significance can be attached to such calculations, but they lend weight to the view that vacancies are involved in this region.

Recently, stress in vacuum deposited films has been measured for a number of metals by Murbach and Wilman (Murbach and Wilman 1953, Wilman 1955). Their measurements were made using films deposited on to copper substrates clamped at one end. The deflection of the free end due to stress in the film was measured with a travelling microscope. Their work involved only the stress at the deposition temperature and they made no annealing experiments. The stress they found in nickel agreed in magnitude with that reported previously (Hoffman, Anders and Crittenden 1953, Hoffman, Daniels and Crittenden 1954) and in copper with that reported here, if allowance is made for the fact that their substrate temperatures rose as much as 100°c during film deposition.

Murbach and Wilman have constructed an alternative theory of stress in vacuum deposited films. Their theory assumes that the stress comes about as a result of contraction of the film as it is depositing. It is assumed that arriving atoms carry sufficient energy to maintain the surface layers at a high temperature. While this temperature exceeds the recrystallization temperature of the metal, the film is able to relax to a stress-free state. When the surface layers then cool to the substrate temperature, the stress develops. Murbach and Wilman therefore conclude that the stress at the deposition temperature should be given by $S=\alpha E(T_r-T_d)$ where E is Young's modulus, α is the temperature coefficient of linear expansion, T_r is the recrystallization temperature and T_d is the deposition temperature. They find that this relationship is obeyed roughly in a number of metals for a given deposition temperature.

The following may be said of the theory as applied to a given metal : (i) it predicts that the initial stress varies linearly with deposition temperature, going to zero when this coincides with the recrystallization temperature; (ii) it includes no mechanism for irreversible effects, but does not exclude the possibility of such effects.

The initial stress in both nickel and copper does vary roughly linearly with deposition temperature. However, it becomes zero in nickel at about 225°c, and in copper at about 75°c. Murbach and Wilman quote recrystallization temperatures as about 600°c in nickel and 200°c in copper. Furthermore, irreversible effects are undeniably present, and are correlated with resistivity changes in films and cold worked materials. Thus the work in copper supports the contention that the intrinsic stress is related to the imperfection density.

ACKNOWLEDGMENTS

The authors wish to thank Mr. T. G. Knorr for his invaluable assistance in the laboratory, Dr. N. S. Rasor for the use of some of his data, and Mr. A. Hruschka for his considerable work in building apparatus.

The authors are also grateful to the Office of Naval Research and the National Carbon Company for financial assistance, and to the Standard Oil Company of Ohio for a fellowship held by one of us (H.S.S.) during the academic year 1954–55.

REFERENCES

BEAMS, J. W., BREAZEALE, J. B., and BART, W. L., 1955, *Phys. Rev.*, **100**, 1657.
BLATT, F. J., 1955, *Phys. Rev.*, **99**, 1708.
BRENNER, A., and SENDEROFF, S., 1949, *Bur. Stand. J. Res., Wash.*, **42**, 105.
BROOM, T., 1954, *Advanc. Phys.*, **3**, 26.
CRITTENDEN, E. C., Jr, and HOFFMAN, R. W., 1953, *Rev. Mod. Phys.*, **25**, 310.
ESHELBY, J. D., 1954, *J. Appl. Phys.*, **25**, 255.
GLEN, J. W., 1955, *Advanc. Phys.*, **4**, 381.
HEAVENS, O. S., 1955, *Optical Properties of Thin Solid Films* (New York : Academic Press), p. 8.
HOFFMAN, R. W., 1949, *Thesis* TM61 (Case Institute of Technology).
HOFFMAN, R. W., ANDERS, F. J., and CRITTENDEN, E. C., Jr, 1953, *J. Appl. Phys.*, **24**, 231.
HOFFMAN, R. W., DANIELS, R. D., and CRITTENDEN, E. C. Jr, 1954, *Proc. Phys. Soc.* B, **67**, **497.**
HOFFMAN, R. W., and STORY, H. S., 1956, *J. Appl. Phys.*, **27**, 193.

JONGENBURGER, P., 1955, *Nature, Lond.*, **175**, 545.

MURBACH, H. P., and WILMAN, H., 1953, *Proc. Phys. Soc.* B, **66**, 905.

OVERHAUSER, A. W., and GORMAN, R. L., 1956, *Phys. Rev.*, **102**, 676.

RASOR, N. S., 1955, O.N.R. *Technical Report* No. 14 (Contract ORI 27303) and 1955, *Phys. Rev.*, **98**, 1555 (Ab).

SANDELL, E. B., 1944, *Colorimetric Determination of Traces of Metals* (New York : Interscience), p. 221.

SONDHEIMER, E. H., 1952, *Advanc. Phys.*, **1**, 1.

TUCKER, C. W., JR, and SAMPSON, J. B., 1954, *Acta Met.*, **2**, 433.

VAN BUREN, H. G., and JONGENBURGER, P., 1955, *Nature, Lond.*, **175**, 544.

WILMAN, H., 1955, *Proc. Phys. Soc.* B, **68**, 474.

Radar Echoes from Inter-Stroke Processes in Lightning

By F. J. HEWITT
National Telecommunications Research Laboratory of the South African Council for Scientific and Industrial Research, Johannesburg, South Africa

Communicated by B. F. J. Schonland ; MS. received 5*th July* 1956 *and in revised form* 26*th February* 1957

Abstract. Radar observations, on a wavelength of 50 cm, of inter-stroke lightning processes provide the following information on the nature of the charged column and the manner in which it is discharged:

(*a*) Inter-stroke junction streamer activity is most intense and most prolonged in the lower regions, that is, at heights ranging from 4–7 km. (*b*) Vertical development of the junction streamers may on occasion take place steadily to a height of about 10 kilometres. The velocities are in good agreement with those derived by Malan and Schonland. The junction streamer activity is somewhat less intense in the upper regions and tends to disappear while activity still persists in the lower regions. (*c*) The lower height limit of junction streamer activity remains at a more or less constant height throughout the flash. (*d*) The horizontal extent of the junction streamer activity in the lower regions is frequently of the order of 1½–2 kilometres. (*e*) The echoing properties of the junction streamers actually decrease during the latter part of the inter-stroke period and a further stroke does not occur until this decrease has taken place. (*f*) A stroke may take place from one or more regions in the charged column, without affecting appreciably the junction streamer process in another region which will nevertheless contribute to a later stroke in the same lightning discharge.

Modifications to the tentative description of the inter-stroke processes given by Malan and Schonland are suggested which would explain these results.

Some information has been gained on the nature and intensity of the noise radiated by lightning at 600 Mc/s. In particular it appears that at this frequency the noise is radiated from comparatively localized sources and may possibly be associated with the stepped and dart leader processes.

---

§ 1. INTRODUCTION

THE use of radar (Hewitt 1953) makes possible the study of certain processes in lightning which cannot easily be observed by other means. Wavelengths in the centimetric region, although attractive from the point of view of resolution and easy availability of equipment, are not satisfactory on account of the background echo intensity from hydrometeors. It has been found that at a wavelength of 50 centimetres radar echoes from lightning can be observed comparatively easily.

According to Malan and Schonland (1951), each stroke in the early stages of a discharge to ground establishes a region of positive charge immediately beneath the remaining portion of the negatively charged vertical column. Immediately after the stroke, the high field existing between this positive charge and the negative column above results in the development of junction streamers. In the inter-stroke period these junction streamers neutralize the positive charge and render the lower portion of the negative column conducting. The new conducting region provides the source of charge for the next stroke.

The radar echoes can be classified into at least two types according to their behaviour. It appears that one type is associated with the inter-stroke junction streamer process and the other with the rapidly decaying ionization in the return stroke channels. The former is the more conspicuous and the easier to observe.

After a brief account of the techniques and procedure used this paper describes the types of record obtained and discusses the observed behaviour of the echoes. Emphasis is placed on the variations in echo intensity with time throughout the flash and on the location of the junction streamers at various stages. The observed behaviour cannot be fitted in all respects to the tentative description of the inter-stroke processes given by Malan and Schonland. A number of modifications are suggested to explain the radar results.

§ 2. EXPERIMENTAL PROCEDURE

The radar equipment used was developed specifically for the purpose. It had the following basic performance characteristics: frequency 600 Mc/s, transmitter power 10–20 kw, pulse length 2 μsec, repetition rate 1000 pulses/sec, aerial gain *c.* 100.

A range–intensity display was used, associated with a continuously rotating drum camera, the range time-bases being at right angles to the direction of motion of the camera film. The beam of the display tube was normally suppressed. On the operation of a trigger sensitive to short bursts of noise picked up by the radar receiving system, the display tube was brightened up for 0·5 sec, resulting in approximately 500 successive radar time traces being recorded separately on the film. Such a trigger system operating on the frequency of the radar equipment and using the same aerial and receiver has the advantage that it is only sensitive to effects occurring within the angular coverage of the radar aerial system. By adjusting the sensitivity, the trigger range may be made roughly equal to that of the radar equipment. While this trigger operated very satisfactorily on the noise associated with the early lightning strokes of a series, it seems probable that it frequently failed to operate on the first stroke. This shortcoming might now be rectified in the light of more detailed knowledge of the waveform of the noise associated with the first as opposed to subsequent strokes.

To obtain height information, an aerial system with two overlapping beams was used. These had their maxima at slightly different angles of elevation. This enabled the angles of elevation of the sources of discrete echoes to be determined by echo amplitude comparison. Two complete receiving channels were provided, with a double beam display tube so connected that the signals received from the higher aerial beam were displayed on the upper trace, and those from the lower aerial beam on the lower trace, with the deflections in opposite directions. The recorded information is discussed in more detail in § 3 where a particular record is discussed.

Although the equipment could detect echoes from lightning up to a range of about 30 km, useful recordings were obtainable up to about half this range only, as the weaker echo components were not observable at the greater ranges. The angular coverage of the equipment was in the region of 20° both horizontally and vertically. The usual procedure followed was to point the aerial system at a region of visually observed lightning activity. When local rain prevented this, the frequency of triggering or of observed radar echoes was used as a guide. The angle of elevation was selected as the best for the study of the junction streamer regions and was not necessarily the best for the study of the echoes from the return stroke channels. Higher performance equipment would be necessary before the vertical coverage could be increased appreciably.

§ 3. THE NATURE OF THE ECHOES

Figure 1 shows in diagrammatic form extracts from the photographic record of the radar echoes from a visually confirmed lightning discharge to ground. The trace-pairs 1, 2, 3, 4 and 5 are radar traces on which any detectable radar echo will appear. The time interval between each successive pair of radar traces is one millisecond. The trace-pairs (S) are subsidiary time traces included to increase the chance of recording short duration noise bursts radiated by the lightning flash. No radar transmissions are associated with these traces, and no short-range radar echo can appear on them. To avoid possible confusion, the range calibration marks 'r' at intervals of 10 microseconds, approximately one mile, were applied only to the subsidiary traces. The range increases from left to right, the first range marker on this record being at three miles.

Examination of the first few subsidiary trace-pairs, 0(S), 1(S) and 2(S) reveals that there is considerably more noise on these traces than on trace-pair 4(S) or the radar trace-pair 5. Similar noise is also seen on the radar traces 1 and 2, though here confusion with radar echoes is possible. This noise indicates the approximate time of occurrence of the stroke, but the process giving rise to it has not yet been positively identified. Radar echoes 'e_1' are clearly discernible on traces 2, 3, 4 and 5, their repetitive nature distinguishing them from the radiated noise which shows no such repetitive properties. Radar echoes are not obvious on trace 1. The recording equipment was triggered a few microseconds after the start of trace 0(S).

Trace-pair 15 occurred 14 milliseconds after trace 1 and shows the echoes immediately before the occurrence of the second stroke. Echo e_1 has decreased somewhat in intensity, whilst two new weak echoes, e_2 and e_3, have appeared. These developed gradually during the inter-stroke period. On trace-pair 16 noise of comparatively high intensity can be seen, associated with the second stroke. This noise is of considerably higher intensity and of much shorter duration than that associated with the first stroke. There is no sign of radiated noise on the trace-pairs 15(S) and 16(S). After the stroke, on trace-pair 17, echoes e_2 and e_3 appear at increased intensity. Echo e_1 is still present, though weak.

The occurrence of noise, followed by the sudden appearance of one or more comparatively strong echoes is a conspicuous feature of all the radar records obtained to date. The most commonly observed echo of this nature persists throughout the inter-stroke period, the intensity frequently decreasing considerably before the next stroke. Such echoes have been tentatively associated with the junction streamer process.

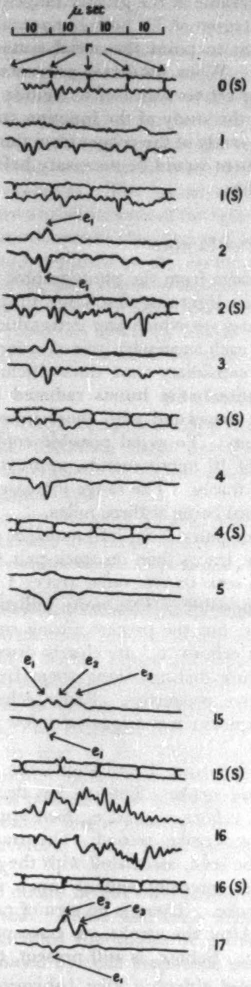


Figure 1. Radar echoes from a lightning discharge to ground.

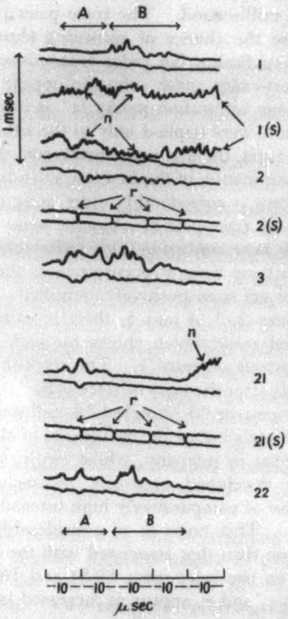


Figure 3. Radar echoes from two successive strokes showing the different effects in the two regions concerned.

§ 4. THE TIME–INTENSITY RELATIONSHIP OF THE JUNCTION STREAMER ECHOES

The time–intensity relationship of echoes associated with the junction streamer process is shown more clearly in figure 2. This relates to one conspicuous echo in another visually confirmed discharge to ground. This echo was the most prominent echo throughout the discharge process and no detectable change in its position could be observed during 250 milliseconds of the record. Information on range only was available when this record was obtained.

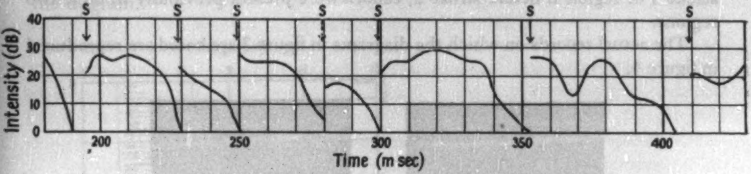


Figure 2. The time–intensity relationship of junction streamer echoes.

The effects during the first 180 milliseconds of this record are not shown, as no obvious trend developed during this stage. The reference level of 0 dB was the internal noise level of the radar receiver. The time of occurrence of the separate strokes is shown by the letter S, and is evidenced on the record by the appearance of noise radiated by some process in the stroke and by a sudden increase in echo intensity. The recovery at 280 milliseconds was not accompanied by detectable radiated noise but is almost certainly evidence of a new though possibly weaker stroke.

Immediately after each stroke the echo intensity is comparatively high. During the first half of each inter-stroke period the echo intensity does not change greatly but in several instances the intensity actually increases slightly with time at some stage. During the second half of each inter-stroke period, the echo intensity decreases steadily and in every case the echo is barely detectable immediately prior to the occurrence of the next stroke.

From records such as this it would appear that the decay in echo intensity and the occurrence of the next stroke are in some way associated.

§ 5. LOCATION OF THE JUNCTION STREAMER ECHOES

In the record on which figure 2 was based, only one conspicuous junction streamer echo was present. This is not always the case. Figure 3 shows in diagrammatic form the echoes immediately before and after the fourth and fifth strokes of a discharge to ground. In this case the fourth stroke results only in changes in the echo intensity in one region, whilst the fifth results only in changes in echo intensity in another region.

The trace-pairs 1, 2 and 3 relate to the fourth stroke, whilst the pairs 21 and 22 relate to the fifth stroke. This record is of the same nature as that illustrated in figure 1 and in fact relates to the same lightning discharge.

It will be seen that on pair 1 there are strong echoes in the region B but not in the region A. Intense noise 'n' radiated by the fourth stroke appears on pair 1(S). On pair 3, strong echoes have appeared in the region A but the echoes in the region B are scarcely, if at all, affected by the stroke. Twenty milliseconds later a similar effect occurs with the fifth stroke except that on this occasion it is the echoes in the region B that are affected. The noise indicating the occurrence of the fifth stroke can be seen on the pair 21. It will also be seen that on pair 21, a few microseconds before the commencement of the noise burst, there are fairly strong echoes in the region A but none in the region B. After the fifth stroke strong echoes appear in the region B as can be seen on pair 22. The echoes in the region A are not appreciably affected.

Although in this example weak echoes are not apparent in region A before stroke 4 or region B before stroke 5, echoes were present previously in these two regions.

The actual records on which the diagrams in figure 3 are based are reproduced in figure 4.

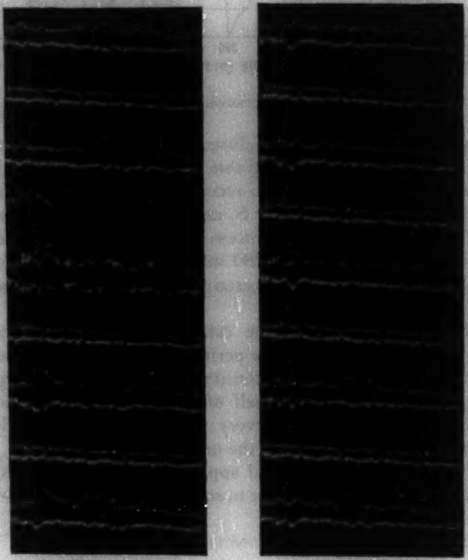


Figure 4. Recorded radar echoes from two successive strokes.

A conspicuous feature of all the records is the fact that, after any stroke other than the first, if echoes appear, they appear in a position previously occupied by weaker echoes at some stage prior to the stroke, within the limit of resolution of the radar equipment. Changes in position of less than about 150 metres would not be observed. Where, on occasion, new echoing regions develop during a lightning discharge, they develop gradually during the inter-stroke period. The next stroke then results in the sudden increase in the echo intensity from some or all of these regions. This gradual development of new echoing regions during the inter-stroke period has been observed to take place both vertically and horizontally.

The development of junction streamer echoes in range throughout a lightning flash to ground is shown diagrammatically in figure 5. This diagram shows the extent of the echoes in range, at various times throughout the record, by means of shaded areas. The thickness of the shading indicates the intensity of the echo at the range and time in question. Each shaded area indicates the state of the echoes averaged over five successive radar traces, i.e. over five milliseconds.

There were at least nine separate strokes discernible on this record, as evidenced by the appearance of noise and the enhancement of certain echoes. The times of occurrence of these strokes are indicated by the numbers 1 to 9 on the diagram.

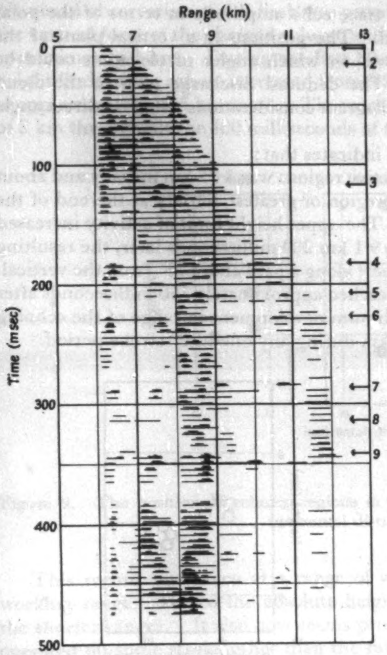


Figure 5. The development of junction streamer echoes in range.

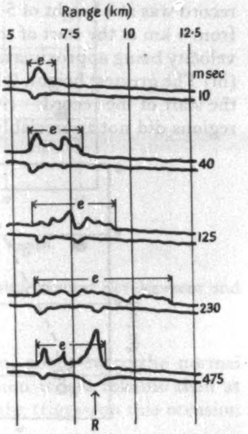


Figure 6. Radar echoes at various stages in a lightning discharge to ground.

It will be seen that during the first 200 milliseconds the progress outwards of the upper range limit of the echoes was remarkably steady. Up to this time echoes appeared on practically every trace on the record. After 200 milliseconds the echoes at long range appeared somewhat less regularly and towards the later stages of the record the long range echoes appeared only spasmodically. At the end of the record (475 milliseconds after the start) there was intense activity towards the lower ranges. In fact, the intensity in this region at the end of the record was greater than that which occurred anywhere else on the record. In particular, it should be noted that the development in slant range took place steadily between strokes and that the occurrence of a new stroke, as indicated by a burst of radiated noise, only enhanced some or all of the echoes previously present.

Figure 6 shows diagrammatically the recorded radar echoes at various stages in this lightning discharge. The extent 'e' in slant range, over which echoes can be observed at these times, is indicated. The intensity of the echo at the range R at 475 milliseconds should also be noted.

The angle of elevation of the sources of some of the echoes recorded can be determined from the upper to lower trace echo amplitude in terms of the polar diagram of the receiver aerial system. The positions in a vertical plane of the sources of the echoes described above, on which angles of elevation could be determined, are shown in figure 7. The deduced discharge path in the cloud is also shown, based on certain polar diagram considerations when no direct angle of elevation could be measured.

Using this information the record indicates that:

(i) The lower height limit of the active regions was 3·75 km initially and about 4 km at the end of the record. The region of greatest activity at the end of the record was at a height of 5·3 km. (ii) The upper height limit of activity increased from 4 km at the start of the record to 9·1 km 200 milliseconds later, the resulting velocity being approximately 25 km sec^{-1} along a path about 25° from the vertical. (iii) The greatest height, 9·8 km, was reached approximately 350 milliseconds after the start of the record. (iv) The maximum development in range of the echoing regions did not appreciably exceed 2 km during any one inter-stroke period.

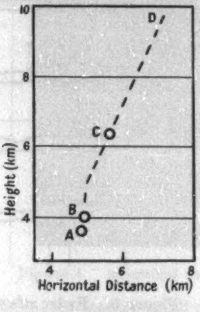


Figure 7.

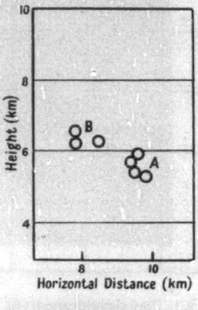


Figure 8.

Figure 7. The position of echoing regions in a flash showing vertical development, A, lowest active region at start of record; B, lowest active region at end of record; D, greatest height reached; ABCD, deduced path of discharge.

Figure 8. The position of echoing regions showing horizontal distribution.

It was not possible to draw any conclusions from this record as to the horizontal development but other records suggest that the activity at a height of 5·3 km at the end of the record might have been due to the tapping of new regions at that height, displaced horizontally from those tapped in the initial stages.

Figure 8 shows the relative positions in a vertical plane of the echo centres in another visually confirmed discharge to ground. Again nine separate strokes were discernible. Throughout the record the more intense echoes were in the region A, at a range of about 10 km from the radar set. The height of this intense echoing region showed a gradual though steady increase from 5·4 to 5·9 km over the whole 0·5 second of the record. Weaker echoes were present spasmodically throughout the record in the region B but there was no evidence of extended vertical development such as was described in the previous record. It will be observed that the activity extends laterally over about 2 km.

These effects were observed to some extent in another record and the relative positions of the echo centres at various times during the flash are shown in figure 9. No visual observation of this flash was possible. Figure 9 illustrates in particular the vertical form of the charged column and the lateral extent in the lower regions, approximately 1 km. During the first 80 milliseconds of this record the upper height limit of the echoes increased from 9·2 to 11·8 km, giving a vertical velocity of approximately 32·5 km sec^{-1}. The most intense activity remained in the region of 8 km throughout the 400 milliseconds of the record.

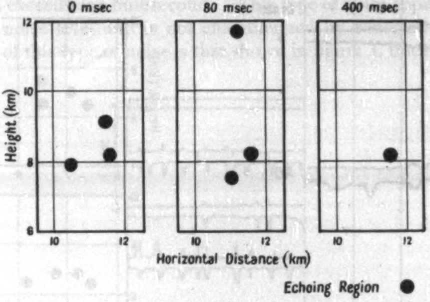


Figure 9. The position of echoing regions in a flash showing vertical development and horizontal distribution.

This record was taken at a range of about 15 km, about twice the normal working range. Hence the absolute height information is less reliable than at the shorter ranges. It also now seems probable that the trigger on this occasion operated on some stroke other than the first.

The behaviour of the echoes may be summarized by stating that, almost without exception, the records so far obtained of the echoes from lightning discharges to ground indicate that junction streamer activity is most intense and most prolonged in the lower regions of the charged column. This intense activity takes place at heights ranging from 4 to 7 km, the lower height limit of the active regions remaining more or less constant throughout each flash. On those occasions when vertical development was observed to take place, the activity was somewhat less intense in the upper regions and tended to disappear while intense activity continued in the lower regions. Activity in the lower regions was observed to extend laterally to about 2 km.

§ 6. Echoes from Air Discharges

The recorded echoes from a visually confirmed air discharge are shown in figure 10 in diagrammatic form at various stages. The development in this case was primarily horizontal. This development is shown at various stages in figure 11. Initially the horizontal extent of the stronger echoes was about 2·5 km, with a very weak echo extending a further 2 km. During the first 40 milliseconds the horizontal extent of the echoes increased to a total of 6·75 km, the development outwards taking place at a velocity of approximately 60 km sec^{-1}. Thereafter the activity persisted in the region in which it was originally observed at greatest strength while the activity, that had developed outward, died away.

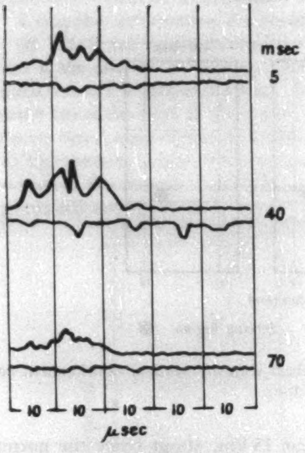


Figure 10. Echoes from an air discharge.

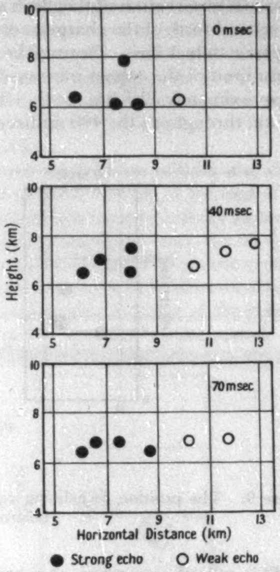


Figure 11. Horizontal development of echoing regions in an air discharge.

The remarkable similarity between echoes from air discharges and those obtained from inter-stroke junction streamer processes in discharges to ground suggests that the mechanism involved in the junction streamer process and in the air discharge may have common features.

§ 7. Echoes from Stroke Channels

Reference has been made in § 1 to an echo component that appears strongly immediately after a stroke but decays in the course of a few milliseconds.

From its behaviour this echo has been associated with the ionization persisting in the highly conducting return stroke channels. It appears that this type of echo occurs at its greatest intensity in the lower regions and that, in general, echoes of this nature in the upper regions are much weaker than those from junction streamer processes. A detailed study has therefore not yet been possible.

§ 8. The Radiated Noise at 600 mc/s

The radiated noise preceding the appearance of echoes has been briefly described previously (Hewitt 1953). More recently it has been possible to observe, with certain limitations, the angle of elevation of the source of noise. On a number of occasions this angle has been observed to increase as the stroke number increases in ground flashes, but there is no obvious correlation between the observed angle of elevation and the location of the sources of the echoes that follow, for, with a display of this type, only short duration samples (about 50 microseconds) at intervals of 500 microseconds are available for study. It appears nevertheless that the noise can come from comparatively localized sources.

The duration of the noise bursts associated with the separate strokes of a ground flash, with the probable exception of the first, is such that the noise frequency appears on two consecutive traces but seldom on three. Thus its duration seldom exceeds one millisecond. This type of noise appears as a general rise in receiver noise level and is not characterized by conspicuous spikes. A typical example of this type of noise is that shown in figure 3, trace-pair 1(S).

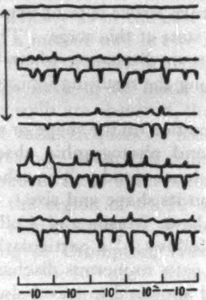


Figure 12. Noise waveform from a cloud flash.

On two occasions the source of the noise appeared to be descending very rapidly, i.e. at about $1{\cdot}5 \times 10^9$ cm sec^{-1}. As the directly observed velocity of the dart leader is somewhat greater than 1×10^9 cm sec^{-1} (Schonland 1938) this together with the observed duration suggests that this noise may be associated with the leader process.

The noise associated with cloud flashes or air discharges is entirely different in form and in duration. The waveform consists of a long series of short individual spikes at intervals of a few microseconds, the whole series lasting for tens of milliseconds. An example of such a waveform is given in figure 12.

Before the flash the noise level was negligible as may be seen from the first trace reproduced. Noise of this nature persisted for over 15 milliseconds.

Radar echoes did not become obvious until some 10–15 milliseconds after the start of the noise.

A few records of visually confirmed flashes to ground show noise of this nature associated with the first recorded stroke. The duration of the noise in such cases was in the region of five milliseconds. It is possible that these were the only occasions on which the trigger operated early enough to record the noise associated with the first stroke. Figure 1 shows what is thought to be first stroke noise approaching this duration, but the intensity is too low to allow the waveform to be studied in detail.

The general power level of the noise associated with both cloud and ground flashes is of the same order. As reported previously the power flux at a range of 8–16 km is at least 10^{-12} wm^{-2} in a pass band of 1 Mc/s centred on 600 Mc/s.

§ 9. THE ORIGIN OF THE ECHOES

The time–intensity relationship of the echoes observed during the inter-stroke period and their location makes it reasonably certain that they are associated with the junction streamer process.

According to Malan and Schonland (1951) each return stroke in the early stages of a discharge to ground establishes a region of positive charge immediately beneath the remaining portion of the negatively charged column. It had been pointed out by Bruce and Golde (1941) that the presence of such a positive charge would give rise to streamers towards the surrounding negatively charged centres. Nothing in the nature of a return stroke is to be expected as the negative column is not in a highly conducting state at this stage. The effect would be somewhat similar to that occurring in an air discharge, where again no return stroke takes place. Currents will flow along the resulting ionized channels until the positive charge has been neutralized.

There is little direct information on the shape or size of these ionized channels within the cloud, as visual and photographic observations are almost always obscured by the surrounding dense cloud. The echoing properties of an ionized channel are very dependent on its shape and size.

On one occasion (Schonland, Malan and Collens 1935) a relatively rare observation was reported as follows: "a particularly active and extraordinarily transparent thunderstorm ... gave numerous discharges within the cloud, all of which branched upwards and appeared to travel slowly upwards from the base. No downward discharges were observed and the impression was that of an upwardly directed discharge exactly similar in nature to the downward stepped leader of a flash to ground, but without a return stroke".

More recently Malan (1956) has drawn attention to some rare photographs of streamers inside the cloud and suggested that "these branching streamers can be considered as demarcating the region of negative charge involved in the discharge to ground".

Malan (1955) has also shown photographically that there is continuous luminosity during the whole of the inter-stroke period, observable only as a glow on account of the presence of cloud except on those occasions such as those referred to above when the actual streamers can be seen.

The conspicuous similarity between the radar echoes from air discharges and from inter-stroke junction streamer processes in discharges to ground has already been pointed out in § 6. In the absence of direct visual information about the discharge these echoes can be distinguished only by the sudden enhancement of echoes after each stroke in the case of a discharge to ground, or by the long series of noise impulses during the early stages of an air discharge. The electrostatic waveforms observed on the ground also suggest that the two processes may have certain common features.

Visual and radar observations therefore suggest that the conducting channels in the junction streamer process are similar in respect of shape and physical size to the other conducting channels in the discharge process, with the exception of the return stroke. The diameters of the highly ionized cores of these various streamer channels appear to be small compared with the wavelength used in these observations. According to Schonland (1937) the diameter of even the main return stroke does not exceed 20 cm, whilst Meek (1939) considers the dart leader tip diameter to be 0·6 cm in his calculation of dart leader velocity, which agrees well with observed velocities. The tortuous nature of the track followed by leader strokes and air discharges is well known.

It is therefore tentatively assumed that the main source of an echo associated with the inter-stroke junction streamer process is a very tortuous, highly ionized trail of diameter small compared with the wavelength used in these observations. The trail may be branched towards the ends, but the stronger echoes can be expected from the denser regions of the trail.

§ 10. THE ELECTRON DENSITY IN JUNCTION STREAMERS

The echoing properties of a trail such as that discussed in § 9 can be calculated in terms of the electron line density as in the case of meteor echoes, though in this case the trail differs from a meteor trail in two important respects. Firstly, the very tortuous nature of the trail renders straight trail interference theory inapplicable, and assumptions have to be made as to the shape. Secondly, at the heights at which meteor trails are observed, the electron collision frequency is considerably lower than the radio frequencies used in their study, while at the heights at which junction streamers occur the collision frequency is very much higher than the frequency used in these observations, e.g. at a height of 90 km the collision frequency is in the region of 6×10^5 sec^{-1} (Nicolet 1953) whereas the radio frequency used in meteor study is of the order of 6×10^7 sec^{-1}. At a height of 7·5 km, according to Crompton, Huxley and Sutton (1953), the electron collision frequency is $2{\cdot}5\times10^{10}$ sec^{-1} when the electron temperature is the same as the gas temperature. In the critical field of 30 kv cm^{-1} necessary to maintain a streamer, the average electron velocity between collisions is $4{\cdot}4\times10^7$ cm sec^{-1} (Schonland and Collens 1933–34). According to Crompton's figures on the dependence of molecular collision cross section on electron velocity, this increased velocity should result in an increase in collision frequency by a factor of 18, giving an electron collision frequency in the streamer of $4{\cdot}5\times10^{11}$ sec^{-1}. This is very considerably greater than the radar frequency of 6×10^8 sec^{-1} used in these observations.

The calculation of the reflecting properties of a thin straight trail of electrons by the method of Blackett and Lovell (1940–41) ignored the effect of collision. This effect, which is not negligible in the case of lightning, may be allowed for by the inclusion of the damping factor $\boldsymbol{m}\nu$ (Appleton and Chapman 1932), in the equation of motion of the electron under the influence of the incident field giving

$$\boldsymbol{m}\frac{d^2x}{dt}+\boldsymbol{m}\nu\frac{dx}{dt}=\boldsymbol{e}E_0\sin\omega t$$

where $\mathbf{e}$ and $\mathbf{m}$ are the charge and mass of the electron respectively, ν the collision frequency and ω the angular radio frequency.

Solution of this equation following the method of Blackett and Lovell gives

$$\rho = \frac{\mathbf{e}^2}{\mathbf{m}r\mathbf{c}^2} \frac{\omega}{(\omega^2+\nu^2)^{1/2}}$$

where ρ is the reflection coefficient of an electron situated at a distance r from the observer.

It will be noted that the inclusion of the damping factor in the original equation has resulted in the introduction of the factor $\omega/(\omega^2+\nu^2)^{1/2}$ in the value of ρ. At a height of 7·5 km the value of $\omega/(\omega^2+\nu^2)^{1/2}$ is approximately 8×10^{-3}, hence the effect of collision is an important one.

If N is the number of free electrons per unit length of the trail, so disposed that the reflected energy from all elements arrives at the radar receiver in phase, then the reflection coefficient of the length l of trail is ρ_l where

$$\rho_l = \frac{\mathbf{e}^2}{\mathbf{m}r\mathbf{c}^2} \frac{\omega}{(\omega^2+\nu^2)^{1/2}} Nl.$$

giving

$$\frac{P_R}{P_T} = \frac{G^2\mathbf{e}^4N^2l^2}{4\mathbf{m}^2r^4\mathbf{c}^2(\omega^2+\nu^2)} \quad \text{and} \quad Nl = \left(\frac{P_R}{P_T}\right)^{1/2} \frac{2\mathbf{m}r^2\mathbf{c}(\omega^2+\nu^2)^{1/2}}{G\mathbf{e}^2}$$

where P_R is the received power at the radar and P_T is the transmitted power. G is the gain of the aerial.

In the simple case of a long straight trail passing through the beam so that a straight line from the radar set, perpendicular to the trail, intersects the trail well within the beam, the echo will be approximately equal to that from the number of electrons, all excited in phase, in a length $(\frac{1}{2}\lambda r)^{1/2}$ of trail. In this case, at a range of 10 km, $(\frac{1}{2}\lambda r)^{1/2} = 5 \times 10^3$ cm and with this value of l the electron line density N becomes $2{\cdot}4 \times 10^{12}$ electrons cm^{-1}.

In the case of a highly tortuous lightning streamer classical interference theory cannot be so easily applied and some assumptions are necessary before even approximate calculations of the reflecting properties can be made.

With a radar pulse length of 2 μsec, all elements of a trail between two concentric spheres centred on the radar set, differing in radius by 300 metres will contribute to one particular echo, as shown in figure 13.

In this zone 300 metres thick, the trail may be considered to consist of a number of relatively short, effectively straight, sections with a number of directional discontinuities. The reflecting property of any one straight section will depend very largely on its direction within the zone. The zone may be considered to be divided into a number of concentric shells, each $\frac{1}{4}\lambda$ or 12·5 cm thick. The echo contribution from the portion of trail in each successive shell will then add or subtract alternatively from the resultant echo, on account of the overall path difference of $\frac{1}{2}\lambda$ introduced between successive shells.

If any straight section cuts the $\frac{1}{4}\lambda$ shells almost perpendicularly to their surfaces, it will pass through a considerable number of shells and, owing to interference effects, the resulting echo will be small, roughly equal to the contribution from the length in any one shell. If it lies almost tangentially to the surface of the shells, the resultant echo may be relatively large as the length intercepted in a shell is much greater. The limiting case is that of the infinitely long straight trail, tangential to the surface of the shell, which was discussed above. In this case the echo will be the same as that from 50 metres of trail all excited in phase. Such an echo contribution would completely swamp the contribution from other straight, less favourably disposed, sections.

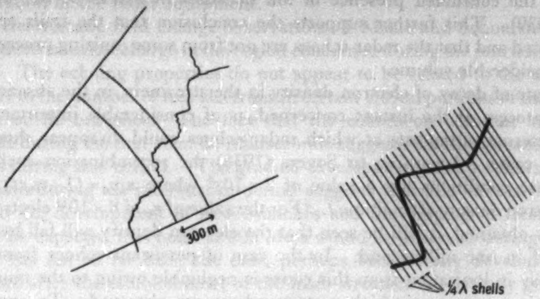


Figure 13. Reflecting properties of a tortuous trail.

From what photographic evidence exists, for example the photograph of the track of an air discharge given by Schonland, Malan and Collens (1935), it would appear that there might be between ten and one hundred such sections in a 300 metre zone. With such a number the existence of a few reasonably favourably disposed sections can be expected, although only occasionally could one be expected to lie completely within a $\frac{1}{4}\lambda$ shell. Thus echoing properties at least approaching those of 50 metres of trail excited in phase may generally be expected, and on occasion it is possible that this intensity will be reached or even exceeded.

The echo amplitude of 10 dB above noise is a typical recorded amplitude at a range of 10 km when echoes are observed over a number of pulse lengths, e.g. 2–3 km. When echo amplitudes of 40 dB or more are recorded, they usually extend over one pulse length only and are due presumably to the less frequently occurring condition when one or more sections of the trail are particularly favourably disposed. There are as yet insufficient data on which to base a statistical study of the amplitude distribution.

The observed echo intensity thus indicates that the electron line density in a junction streamer or an air discharge channel is in excess of $2{\cdot}4 \times 10^{12}$ electrons cm^{-1}. With a trail diameter of 2 cm the minimum electron density within the trail becomes 8×10^{11} electrons cm^{-3}.

The value of the line density can also be calculated, on the basis of a single streamer, from the magnitude of the positive charge $+q$ that is neutralized during the inter-stroke interval and the electron drift velocity. For early strokes, Malan and Schonland (1951) observed that $+q$ ranges from $Q/2{\cdot}5$ to $Q/1{\cdot}03$ where Q is the charge carried to ground by the stroke; $+q$ is thus of the order of one coulomb. To discharge this in from 30 to 50 milliseconds requires a current in the region of 25 amperes, or $10^{19} \times 25/1{\cdot}6$ electrons sec^{-1}. The minimum electron drift velocity in a lightning streamer appears to be determined by the electron mean free path and the critical field strength below which ionization by collision will not occur. Schonland (1938) shows that the calculated value is close to the observed value of about 2×10^{7} cm sec^{-1}. Applying this value, the electron line density in the junction streamer becomes $7{\cdot}8 \times 10^{12}$ electrons cm^{-1}. With a streamer diameter of 2 cm the electron density within the streamer becomes $2{\cdot}5 \times 10^{12}$ electrons cm^{-3}.

The luminosity observed by Malan (1955) throughout the inter-stroke interval, indicates the continued presence of ion densities in excess of 10^{12} ions cm^{-3} (Meek 1939). This further supports the conclusion that the trails are highly concentrated and that the radar echoes are not from some ionizing process spread over a considerable volume.

The rate of decay of electron density in the streamers, in the absence of any ionizing process at the instant concerned, is of considerable importance, as it largely determines the rate at which radar echoes could disappear should such processes cease. According to Sayers (1938) the recombination coefficient α under these conditions has a value of 2×10^{-6} where $n/n_0 = (1 + n_0\alpha t)^{-1}$ and n is the electron density after a time t. For the value of n_0 of 8×10^{11} electrons cm^{-3} previously obtained, it will be seen that the electron density will fall by a factor of 6×10^{-4} in one millisecond. In the case of persistent echoes from meteor trails, decay in ionization from this cause is negligible owing to the much lower atmospheric pressure at which meteor echoes are observed. To explain the persistence of lightning echoes over many milliseconds on a similar bais would require extremely high initial electron densities for which there appears to be no justification. The persistence of the lighning echoes thus suggests the existence of continuous ionizing processes during the time in question. In the case of a junction streamer carrying a current, a continuous ionizing process along the length of the streamer is essential to its existence, and its origin is the potential drop along the streamer necessary for the maintenance of the current. It has therefore been concluded that with a radar recurrence time of one millisecond the presence of recurrent echoes indicates the existence of ionizing processes at the time in question.

Observations of Biondi and Brown (1949), although not directly applicable to these conditions, enable a lower limit to be set to the maximum rate of decay which confirms the above conclusion.

§ 11. THE DEVELOPMENT OF JUNCTION STREAMERS

The more important characteristics of the junction streamer echoes are as follows:

(*a*) The amplitude of the predominant junction streamer echo is comparatively large at the commencement of the inter-stroke interval. It then remains more or less constant or may even increase during the first half of the inter-stroke interval. During the last 20 milliseconds or so before the next stroke from the region in question, it invariably decreases steadily by about 20 dB or more.

(*b*) The predominant junction streamer echo remains in the same place throughout the inter-stroke interval, within the resolution of the equipment, and the greatest activity frequently persists throughout the discharge in much the same region.

(*c*) When new echoing regions develop during the discharge, they build up gradually from the threshold of detectability during the inter-stroke interval, the next stroke resulting in the sudden enhancement of some or all of the echoes.

The fact that the predominant echo is large at the commencement of the inter-stroke interval indicates that the intense portion of the junction streamer develops almost fully within one millisecond of the preceding return stroke. This rapid development is to be expected from the streamer tip velocity in the region of 10^7 cm sec^{-1} and the fact that the distance involved in so far as the most intense portion of the streamer is concerned does not appear to exceed 150 metres, the resolution of the radar equipment.

The electrostatic field change observations of Malan and Schonland indicate a slow reduction in net charge in the region concerned during the junction streamer process. The echoing properties do not appear to be related to the net charge but rather to the number of free electrons in certain critical portions of the junction streamer or its branches. The fact that the echoes on occasion increase in amplitude during the first 10 or 20 milliseconds suggests that the streamer current increases during this period. The gradual development of new echoes observed on occasion during the inter-stroke period suggests that this increase in current is due to the development of new branches and the tapping of fresh regions. It might be expected that noise at 600 Mc/s would be observed during this stage but this has not so far been the case, perhaps because any development that takes place after the establishment of the main streamer is relatively gradual and weak.

The predominant echo, and hence presumably the most intense portion of the junction streamer, does not move appreciably during the inter-stroke period, though the weaker echoes, that may develop gradually, spread away from the main echo in the general direction of the rest of the negative column. This development may extend over several kilometres, indicating the extent of the negative region tapped during the inter-stroke period. The fact that junction streamer activity persists in greatest strength in much the same region during successive inter-stroke periods suggests that the greater part of the positive charge brought up by each return stroke occupies much the same position throughout the discharge.

§ 12. The Time of Occurrence of a Stroke

The relation between the decreasing junction streamer echo intensity and the time of occurrence of the next stroke suggests that the amount of positive charge still remaining in the region after a stroke to ground, may be the main factor determining the time of occurrence of the next stroke. If the echo intensity is an indication of the junction streamer current, the radar results indicate that the current decreases considerably in the 10–20 milliseconds immediately prior to a stroke. This could be due to the gradual falling off of the potential gradient along the streamer resulting from the steady neutralization of the positive charge. According to Malan (1956), ionization persists in the main stroke channel for 100 milliseconds or more after the return stroke. During the inter-stroke period this channel will be shielded from the negative region by the positive charge, but the shielding effect will decrease as the positive charge is neutralized and a value will eventually be reached when a new leader stroke can take place.

The return stroke restores the positive charge, re-establishing the local fields and the junction streamers. The shielding effect on the still ionized main stroke channel will also be restored.

§ 13. The Distribution of the Negative Charge in the Cloud

The processes suggested in §§ 11 and 12 could also explain the occurrence of a stroke which results in changes in echo intensity from certain regions of the cloud but not from others. This may be due to the spatial distribution of the negative charges, and supports the observations of Malan (1955) that the distribution of negative charge in the cloud is not uniform.

A record showing this effect very clearly was described in §3. Figure 14 illustrates the effect diagrammatically, with junction streamers 'j' in the two regions A and B at the same instant. If, for some reason, the activity at A were more advanced in time than that at B, as appears to have been the case before the fourth stroke in the record described in §3, the positive charge C_a shielding the region A from the weakly ionized stroke channel 'c', will reach the minimum critical value while there is still considerable positive charge C_b shielding the region B from the stroke channel. Immediately prior to this stage, radar echoes from the streamers in region A will be very weak owing to the weak junction streamer activity consequent on the diminished positive charge C_a. The echoes will still be strong from the region B where intense junction streamer activity is maintained as a result of the charge C_b still being comparatively strong.

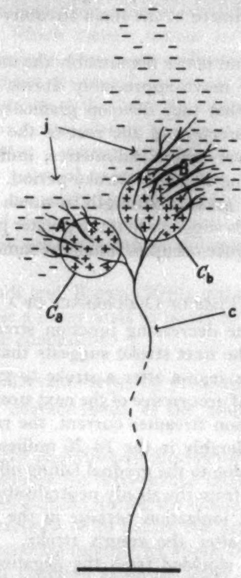


Figure 14. Junction streamer activity in two regions at the same time, only one region contributing to a particular stroke.

When the positive charge C_a reaches the minimum critical value, a stroke will occur from the region A but the region B will not contribute owing to the continued existence of the charge C_b. This stroke corresponds to the fourth stroke in the record previously described. The stroke process from A will restore the positive charge in the region A, resulting in renewed junction streamer activity and hence renewed radar echoes.

After the occurrence of the stroke from the region A, the junction streamer activity in the region B will be more advanced in time than that at A. It is therefore to be expected that the next stroke will occur from the region B. In the record described in § 3 examination of the radar trace-pairs associated with the fifth stroke confirms that this did in fact occur.

ACKNOWLEDGMENTS

The author wishes to express his thanks to the South African Council for Scientific and Industrial Research for permission to publish this paper and to Dr. B. F. J. Schonland, who suggested this research, for his assistance and advice.

REFERENCES

APPLETON, E., and CHAPMAN, S., 1932, *Proc. Phys. Soc.*, **44**, 246.
BIONDI, M. A., and BROWN, S. C., 1949, *Phys. Rev.*, **76**, 1697.
BLACKETT, P. M. S., and LOVELL, A. C. B., 1940–41, *Proc. Roy. Soc.* A, **177**, 183.
BRUCE, C. E. R., and GOLDE, R. H., 1941, *J. Instn. Elect. Engrs.*, **88**, 498.
CROMPTON, R. W., HUXLEY, L. G. H., and SUTTON, D. J., 1953, *Proc. Roy. Soc.* A, **218**, 507.
HEWITT, F. J., 1953, *Proc. Phys. Soc.* B, **66**, 895.
MALAN, D. J., 1955, *Ann. Géophys., II, Fas.*, **4**, ; 1956, *Geofis. Pura Appl.*, **34**, 221, 224.
MALAN, D. J., and SCHONLAND, B. F. J., 1951, *Proc. Roy. Soc.* A, **206**, 145.
MEEK, J. M., 1939, *Phys. Rev.*, **55**, 972.
NICOLET, M., 1953, *J. Atmos. Terr. Phys.*, **3**, 200.
SAYERS, J., 1938, *Proc. Roy. Soc.* A, 169, 83.
SCHONLAND, B. F. J., 1937, *Phil. Mag.*, **23**, 503 ; 1938, *Proc. Roy. Soc.* A, **164**, 133 ; 1951, *Ibid.*, **206**, 145.
SCHONLAND, B. F. J., and COLLENS, H., 1933–34, *Proc. Roy. Soc.* A, **143**, 654.
SCHONLAND, B. F. J., MALAN, D. J., and COLLENS, H., 1935, *Proc. Roy. Soc.* A, **152**, 595.

Nuclear Magnetic Resonance Spectrum for Isosceles Triangular Configurations of Nuclei

By E. R. ANDREW and N. D. FINCH
Physics Department, University College of North Wales, Bangor, Caernarvonshire

MS. received 8th *May* 1957

Abstract. The nuclear magnetic resonance spectrum is calculated for an assembly of isotropically oriented systems of three identical nuclei of spin number $\frac{1}{2}$ situated at the corners of an isosceles triangle. The theoretical spectra for triangles departing in various degrees from equilateral shape are compared with those recorded by Richards and Smith for the monohydrates of nitric and perchloric acids in polycrystalline form. It is concluded that if the oxonium ion in these acid hydrates is not equilateral, it does not depart from that shape by more than 10%. Interionic broadening is examined and it is concluded that for the acid hydrates a Gaussian function gives a good description of this broadening.

§ 1. INTRODUCTION

THE magnetic resonance spectrum of nuclei of spin number $\frac{1}{2}$ in diamagnetic crystals is predominantly determined by the magnetic dipolar interactions between the nuclei. Because these interactions are proportional to the inverse cubes of the separating distances, the shape of the spectrum is mainly determined by the interactions between near neighbours in the crystal lattice. A crystalline specimen containing relatively isolated identical groups of nuclei therefore yields a spectrum characteristic of the group. From the spectrum the structure of the group can often be determined, and in some cases the positions of the nuclei in the unit cell can be located.

A number of materials contain nuclei with spin number $\frac{1}{2}$ clustered in relatively isolated groups of three. Examples are organic compounds containing methyl groups (CH_3) and trifluoromethyl groups (CF_3), acid hydrates containing oxonium ions (H_3O^+), and ammines containing ammonia molecules (NH_3). The theory of the spectrum characteristic of relatively isolated identical groups of three nuclei, similarly oriented in a single crystal, has been given by Andrew and Bersohn (1950); this earlier paper will be referred to as I. The spectrum was found to consist of seven lines, namely a centre line and three symmetrically placed pairs of satellites. It is not always possible to obtain a single crystal for measurement and experimenters have so far used polycrystalline specimens. In I the spectrum was calculated for polycrystalline material for the particular case of groups of three nuclei lying on the corners of an equilateral triangle. The spectrum for a given crystal grain depends markedly upon the orientation of the grain relative to the applied magnetic field. It was therefore necessary to integrate the contributions to the spectrum from grains of all orientations.

In this paper the spectrum is calculated for polycrystalline specimens containing groups of three identical nuclei on the corners of identical isosceles triangles. Here the spectrum for a given crystal grain depends more intricately upon the orientation of the grain relative to the applied magnetic field, and the integration of contributions from grains of all orientations is more complicated. The results are applied in § 4 to the spectra of some polycrystalline acid hydrates recorded by Richards and Smith (1951).

§ 2. CALCULATION OF THE SPECTRUM

2.1. *Spectrum for a Single Crystal*

Consider three identical nuclei of magnetic moment μ and spin number $\frac{1}{2}$ arranged at the corners of any general triangle whose sides are r_{12}, r_{23} and r_{31}. Let a steady magnetic field **H** be applied in a direction making angles θ_{ij} with the three sides r_{ij}. Let H^* be the resonant field strength for the free nuclei at the radiofrequency in use, and let $h=H-H^*$. It was shown in I that the absorption spectrum for a large number of relatively isolated triangular groups of this kind, similarly oriented, consists of a central line at $h=0$, and three pairs of satellite lines at values of h given by $\pm y/\mu$, $\pm(3x+y)/2\mu$, $\pm(3x-y)/2\mu$, where

$$\left.\begin{array}{lll} & x=\tfrac{1}{4}(A_{12}+A_{23}+A_{31}), & y^2=x^2+a^2+b^2 \\ & a=(\sqrt{6}/4)(A_{31}-A_{23}), & b=(\sqrt{2}/4)(2A_{12}-A_{23}-A_{31}) \\ \text{and} & \multicolumn{2}{c}{A_{ij}=2\mu^2 r_{ij}^{-3}(3\cos^2\theta_{ij}-1).} \end{array}\right\} \quad \ldots\ldots(1)$$

The normalized intensities of the seven lines are as follows:

$$\left.\begin{array}{llll} \text{Position of line, } h & 0 & \pm y/\mu & \pm(3x\pm y)/2\mu \\ \text{Normalized intensity, } p & \dfrac{1}{8}\left(1+\dfrac{3x^2}{y^2}\right) & \dfrac{3}{16}\left(1-\dfrac{x^2}{y^2}\right) & \dfrac{1}{8}\left(1\pm\dfrac{x}{y}\right). \end{array}\right\} \quad \ldots\ldots(2)$$

Consider now the isosceles triangle shown in figure 1. Let the two equal sides be $r_{23}=r_{31}=P$, let the third side $r_{12}=R$, and let $\lambda=P/R$. Let the steady field **H** be applied in a direction making an angle ψ with the normal to the plane of the triangle and let the azimuth of **H** in the plane of the triangle relative to the side

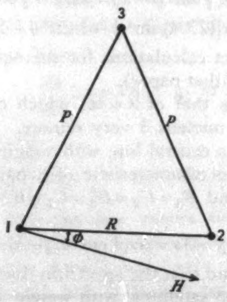


Figure 1. Isosceles triangle. The applied field **H** is inclined at an angle ψ to the normal to the plane of the triangle, and its azimuth in this plane is ϕ relative to the side R.

r_{12} be ϕ. Let θ_{ij} be the angles between the sides r_{ij} and $\mathbf{H}$. Then we readily find that

$$\left.\begin{aligned}\cos\theta_{12}&=\sin\psi\sin\phi\\ \cos\theta_{23}&=\sin\psi[\sin\phi-\sqrt{(4\lambda^2-1)}\cos\phi]/2\lambda\\ \cos\theta_{31}&=\sin\psi[\sin\phi+\sqrt{(4\lambda^2-1)}\cos\phi]/2\lambda.\end{aligned}\right\}\qquad\ldots\ldots(3)$$

Substituting these expressions for the $\cos\theta_{ij}$ in the general equations (1), we find that the values of x and y for the isosceles triangle are given by

$$x=\mu\alpha(B_1\sin^2\psi\sin^2\phi+B_2\sin^2\psi-B_3),\qquad\ldots\ldots(4)$$

$$\begin{aligned}y=\mu\alpha(&C_1\sin^4\psi\sin^4\phi-C_2\sin^4\psi\sin^2\phi+C_3\sin^4\psi\\ &-C_4\sin^2\psi\sin^2\phi-C_5\sin^2\psi+C_6)^{1/2}\end{aligned}\qquad\ldots\ldots(5)$$

where $\alpha\equiv\frac{3}{2}\mu R^{-3}$, and the coefficients B and C, which are functions solely of λ and therefore constant for a given triangle, are given by

$$\left.\begin{aligned}B_1&=1-2/\lambda^3+1/\lambda^5,\\ B_2&=2/\lambda^3-1/2\lambda^5,\\ B_3&=\tfrac{1}{3}(1+2/\lambda^3),\\ C_1&=3(3+4/\lambda^3-2/\lambda^5+4/\lambda^6-12/\lambda^8+3/\lambda^{10}),\\ C_2&=(3/\lambda^3)\,(4-1/\lambda^2+8/\lambda^3-14/\lambda^5+3/\lambda^7),\\ C_3&=(12/\lambda^{10})\,(\lambda^2-\tfrac{1}{4})^2,\\ C_4&=6-2/\lambda^5-8/\lambda^6+4/\lambda^8,\\ C_5&=(4/\lambda^5)\,(2/\lambda^3-1)\,(\lambda^2-\tfrac{1}{4}),\\ C_6&=1-4/(3\lambda^3)+4/(3\lambda^6).\end{aligned}\right\}\qquad\ldots\ldots(6)$$

The spectrum for a single crystal containing relatively isolated identical isosceles triangular groups of nuclei is therefore given by (2) with (4), (5) and (6). The positions and intensities of the seven lines are complicated functions of the polar and azimuth angles ψ and ϕ; the spectrum will therefore be highly anisotropic.

An instructive check on equations (4), (5) and (6) is provided by a consideration of two special cases. An equilateral triangle is a special isosceles triangle having a ratio $\lambda=1$. We then find $B_1=C_1=C_2=C_4=0$ and

$$x=\mu\alpha(\tfrac{3}{2}\sin^2\psi-1)=\mu\alpha(\tfrac{1}{2}-\tfrac{3}{2}\cos^2\psi),\qquad\ldots\ldots(7)$$

$$y=\mu\alpha[(27/4)\sin^4\psi-3\sin^2\psi+1]^{1/2},\qquad\ldots\ldots(8)$$

in agreement with the direct calculations for the equilateral triangle given in I (see equations (2) and (3) of that paper).

A second special case is that of $\lambda=\infty$, which corresponds to a close pair of nuclei 1 and 2 with the nucleus 3 very remote. The spectrum in this case should therefore consist of a central line with weight $\frac{1}{3}$ due to the lone nucleus superposed on the spectrum characteristic of a pair of nuclei with weight $\frac{2}{3}$. Putting $\lambda=\infty$ in (6), we find $B_2=C_2=C_3=C_5=0$, and with (4) and (5) using (3) we get

$$y=3x=\mu\alpha(3\cos^2\theta_{12}-1).\qquad\ldots\ldots(9)$$

Substituting (9) in (2) we find that the spectrum does indeed consist of a central line, with weight $\frac{1}{3}$, and two satellites, with weight $\frac{1}{3}$ each, whose positions are given by

$$h=\pm\alpha(3\cos^2\theta_{12}-1)=\pm\tfrac{3}{2}\mu R^{-3}(3\cos^2\theta_{12}-1).\qquad\ldots\ldots(10)$$

These positions of the two satellites are in agreement with the direct calculation of Pake (1948) for pairs of identical nuclei.

2.2. *Spectrum for Polycrystalline Material*

Let us now calculate the spectrum for relatively isolated isosceles triangular groups isotropically oriented in polycrystalline material. The spectrum will be a summation of the widely differing spectra of the individual crystal grains. First consider those triangles whose polar angles lie in the interval ψ, $\psi+d\psi$; this group of triangles, which we will call group ψ, constitutes a fraction $\sin\psi d\psi$ of the whole isotropic assembly. Among this group ψ let us consider those triangles whose azimuth angle lies in the interval ϕ, $\phi+d\phi$; this sub-group constitutes a fraction $d\phi/2\pi$ of group ψ. The triangles in this sub-group all have the same orientation ψ, ϕ and give the characteristic spectrum of centre line and three pairs of satellites. Consider the kth line of this spectrum. Let its normalized intensity, given by (2), be $p_k(\psi, \phi)$. The intensity which this line contributes to the normalized spectrum of group ψ is therefore $p_k(\psi, \phi)d\phi/2\pi$. If we describe the spectrum from all crystal grains in group ψ by a normalized shape function $f(\psi, h)$, then

$$f(\psi, h)=\frac{1}{2\pi}\sum_k p_k(\psi, \phi)\left|\frac{d\phi}{dh}\right|_k. \qquad \ldots\ldots(11)$$

The spectrum of the whole isotropic assembly may be described by a normalized shape function $F(h)$, which will be given by an integration of $f(\psi, h)$ over all groups ψ. We get

$$F(h)=\int_0^{\pi/2} f(\psi, h)\sin\psi\, d\psi. \qquad \ldots\ldots(12)$$

The first task is to find values of $(d\phi/dh)_k$ for each line, so that values of $f(\psi, h)$ may be obtained from (11). Consider first the line $h=y/\mu$. It is convenient to write ξ for $\sin^2\phi$ in the expression (5) for y, giving

$$h/\alpha=y/\mu\alpha=[D_1\xi^2-D_2\xi+D_3]^{1/2}, \qquad \ldots\ldots(5')$$

where

$$\left.\begin{aligned} D_1&=C_1\sin^4\psi \\ D_2&=C_2\sin^4\psi+C_4\sin^2\psi \\ D_3&=C_3\sin^4\psi-C_5\sin^2\psi+C_6. \end{aligned}\right\} \qquad \ldots\ldots(13)$$

From the definition of ξ we have

$$\frac{d\phi}{dh}=\tfrac{1}{2}\xi^{-1/2}(1-\xi)^{-1/2}\frac{d\xi}{dh}, \qquad \ldots\ldots(14)$$

while from (5′) ξ is given by

$$\xi=\frac{1}{2D_1}\left[D_2\pm\sqrt{\left\{D_2^2-4D_1\left(D_3-\frac{h^2}{\alpha^2}\right)\right\}}\right]. \qquad \ldots\ldots(15)$$

From (5′) we see that the line moves from $h=D_3^{1/2}\alpha$ when $\phi=0$ and $\xi=0$, to $h=(D_3-D_2^2/4D_1)^{1/2}\alpha$ at a turning point $\xi_t=D_2/2D_1$, and back to $h=(D_1-D_2+D_3)^{1/2}\alpha$ when $\phi=\pi/2$ and $\xi=1$; in the latter part of the range the plus sign is taken in (15), and in the first part the minus sign is taken. The sequence is reversed in the next quadrant from $\phi=\pi/2$ to π; the behaviour in the remaining two quadrants is the same as in the first two. When $D_2\geqslant 2D_1$

there is no turning point within each quadrant of azimuth, and the line moves monotonically from $h = D_3^{1/2}\alpha$ when $\phi = 0$ and $\xi = 0$ to $h = (D_1 - D_2 + D_3)^{1/2}\alpha$ when $\phi = \pi/2$ and $\xi = 1$; from (13) we see that the condition $D_2 \geqslant 2D_1$ for the absence of a turning point is that $\psi > \sin^{-1}[C_4/(2C_1 - C_2)]^{1/2}$.

Thus to evaluate the contribution of the line $h = y/\mu$ to the shape function $f(\psi, h)$ for a particular value of ψ we calculate the terms D from (13) for that ψ and derive values of ξ and $d\xi/dh$ for a suitable value of h in the range just described. The values of ξ and $d\xi/dh$ are substituted in (14) to obtain values of $d\phi/dh$. The ratio x/y is found using (4) and (5) enabling the corresponding values of $p(\psi, \phi)$ to be calculated from (2). The contribution of this line to $f(\psi, h)$ is then found using (11).

The behaviour of the lines $h = (3x \pm y)/2\mu$ is similar to that for $h = y/\mu$. From its value at $\phi = 0$, $\xi = 0$, h traverses a range of values to a turning point ξ_t and then moves back to the value for $\phi = \pi/2$, $\xi = 1$. The sequence is reversed in the next quadrant of azimuth and the behaviour in the remaining two quadrants is the same as in the first two. The turning point ξ_t is given by the expression $[D_2 \mp \surd(D_2^2 - 4D_1D_6)]/2D_1$ where $D_6 = (D_4^2 D_3 - D_2^2)/4(D_4^2 - D_1)$, $D_4 = 3B_1 \sin^2\psi$, and D_1, D_2 and D_3 are defined in (13); for the line $h = (3x + y)/2\mu$ the minus sign is to be taken in ξ_t and for the line $h = (3x - y)/2\mu$ the plus sign is to be taken.

In order to find $d\phi/dh$ for the lines $h = (3x \pm y)/2\mu$ using (14), we must express ξ as a function of h. Using (4) and (5) we find after rearrangement that

$$\xi \equiv \sin^2\phi = \frac{1}{2E_1}[-E_2 \pm \surd(E_2^2 - 4E_1E_3)], \qquad \ldots\ldots(16)$$

where

$$\left.\begin{aligned} E_1 &= (9B_1^2 - C_1)\sin^4\psi \\ E_2 &= (18B_1B_2 + C_2)\sin^4\psi - \left(18B_1B_3 - C_4 + 12B_1\frac{h}{\alpha}\right)\sin^2\psi \\ E_3 &= (9B_2^2 - C_3)\sin^4\psi + \left(C_5 - 18B_2B_3 - 12B_2\frac{h}{\alpha}\right)\sin^2\psi \\ &\qquad + \left(9B_3^2 - C_6 + 12B_3\frac{h}{\alpha} + 4\frac{h^2}{\alpha^2}\right). \end{aligned}\right\} \qquad \ldots\ldots(17)$$

The remaining satellite lines $h = -y/\mu$, $h = -(3x \pm y)/2\mu$ are symmetrically placed about the centre line relative to the three lines we have so far considered. Their contribution to $f(\psi, h)$ is therefore given by replacing h by $-h$.

Finally we must consider the contribution to $f(\psi, h)$ of the centre line from all triangles in group ψ. Its intensity is

$$\frac{1}{2\pi}\int_0^{2\pi} p(\psi, \phi)\,d\phi = \frac{1}{16\pi}\int_0^{2\pi}\left(1 + \frac{3x^2}{y^2}\right)d\phi, \qquad \ldots\ldots(18)$$

using (2).

Having evaluated $f(\psi, h)$ for various ψ we can now calculate $F(h)$ from (12), performing the integration numerically. The integrated contribution of the centre line is considered in more detail at the end of this section.

We have carried out the calculations for $\lambda = 1{\cdot}1$ and $1{\cdot}2$ obtaining the results given in the table. Values for $\lambda = 1$ (equilateral triangle) are also given. Values of $F(h)$ are given at intervals of $0{\cdot}1\alpha$; each value quoted is the mean of $F(h)$ over the range $h \pm 0{\cdot}05\alpha$. Careful attention was paid to the ends of the ranges of h covered by each component line, since $d\phi/dh$ becomes infinite at these points.

The calculations may be readily extended to other values of λ though this involves much numerical work.

Values of $\alpha F(h)$.

h/α	$\lambda=1$	$\lambda=1{\cdot}1$	$\lambda=1{\cdot}2$	h/α	$\lambda=1$	$\lambda=1{\cdot}1$	$\lambda=1{\cdot}2$
0	0·037	0·061	0·105	1·2	0·186	0·254	0·287
0·1	0·037	0·054	0·112	1·3	0·189	0·149	0·197
0·2	0·037	0·086	0·124	1·4	0·193	0·467	0·245
0·3	0·132	0·188	0·193	1·5	0·202	0·256	0·135
0·4	0·254	0·272	0·208	1·6	0·215	0·242	0·114
0·5	0·162	0·166	0·283	1·7	0·248	0·180	0·128
0·6	0·139	0·172	0·260	1·8	0·400	0·111	0·075
0·7	0·128	0·204	0·268	1·9	0·157	0·119	0·041
0·8	0·185	0·344	0·258	2·0	0·161	0·031	0.026
0·9	0·209	0·163	0·311	2·1	0·215	0·052	0·041
1·0	0·189	0·156	0·353	2·2	0·195	0·017	0·013
1·1	0·187	0·362	0·325	2·3	0·000	0·000	0·000
				Centre line	0·191	0·189	0·185

We notice from (2) that the combined intensities of the four satellite lines $h=\pm(3x\pm y)/2\mu$ is always $\frac{1}{2}$, for all triangle shapes and orientations. The integrated contribution of these four lines to the shape function $F(h)$ must therefore be $\frac{1}{2}$, and this provides a useful check on the numerical work.

A further check of the data in the table is provided by calculating the second moment (mean square width) of the spectrum, defined as

$$S_2=\int_{-\infty}^{\infty} h^2 F(h)dh. \qquad \ldots\ldots(19)$$

Van Vleck (1948) has provided formulae for the second moment of dipolar-broadened spectral lines, which when applied to our problem gives

$$S_2=\frac{8}{15}\alpha^2\left(1+\frac{2}{\lambda^6}\right). \qquad \ldots\ldots(20)$$

The second moments evaluated from the table for the three values of λ, using (19), agree with those calculated from (20) within a few tenths of a per cent.

We now take up again the integrated contribution of the centre line to $F(h)$, which is obtained by integrating (18) over all groups ψ. We get

$$\frac{1}{16\pi}\int_0^{2\pi} d\phi\int_0^{\pi/2}\left(1+\frac{3x^2}{y^2}\right)\sin\psi\, d\psi=\frac{1}{8}+\frac{3}{16\pi}\int_0^{2\pi} d\phi\int_0^{\pi/2}\frac{x^2}{y^2}\sin\psi\, d\psi. \qquad \ldots\ldots(21)$$

Using expressions (4) and (5) for x and y we find eventually that

$$\int_0^{\pi/2}\frac{x^2}{y^2}\sin\psi\, d\psi=\frac{K_1^2}{K_2}+\frac{K_1^2K_4}{4K_2u}\left(\frac{K_5}{K_7}-1\right)\ln\left(\frac{1+u+K_7}{1-u+K_7}\right)$$
$$+\frac{K_1^2K_4}{4K_2}\left(\frac{K_5}{K_7}+1\right)\left(K_7-\frac{u^2}{4}\right)^{-1/2}\left[\tan^{-1}\left\{\frac{1+\frac{1}{2}u}{(K_7-\frac{1}{4}u^2)^{1/2}}\right\}+\tan^{-1}\left\{\frac{1-\frac{1}{2}u}{(K_7-\frac{1}{4}u^2)^{1/2}}\right\}\right], \qquad \ldots\ldots(22)$$

where

$K_1=B_1\sin^2\phi+B_2$ $\qquad$ $K_5=[(1-B_3/K_1)^2-K_7^2]/K_4$

$K_2=C_1\sin^4\phi-C_2\sin^2\phi+C_3$ $\qquad$ $K_6=K_3/K_2-2$

$K_3=C_4\sin^2\phi+C_5$ $\qquad$ $K_7=[1-(C_6-K_3)/K_2]^{1/2}$

$K_4=3B_3/K_1-K_3/K_2$ $\qquad$ $u^2=2K_7-K_6.$

Expression (22) is evaluated for various ϕ and substituted in (21) to give by numerical integration the normalized intensity of the centre line. This is plotted in figure 2 for the whole range of λ from 0·5 to ∞; when $\lambda = 0{\cdot}5$ the three nuclei are collinear, while when $\lambda = \infty$ we have a pair of nuclei and one remote nucleus. The intensity is 0·284 for $\lambda = 0{\cdot}5$ and falls through a shallow minimum rising to 0·191 for $\lambda = 1$ after which it falls again to 0·167 (i.e. 1/6) for $\lambda = \infty$. This might at first sight appear to contradict the requirement, discussed earlier, that the centre line shall have intensity $\frac{1}{6}$ when $\lambda = \infty$. However, the intensity which is apparently missing comes from the satellites $h = \pm(3x - y)/2\mu$ which coalesce to the centre with intensity 1/12 each as $\lambda \to \infty$ (see equation (9)). Thus although the centre line proper declines slightly in intensity with increasing λ the central part of the spectrum actually becomes increasingly enhanced on account of the movement inwards of this pair of satellites.

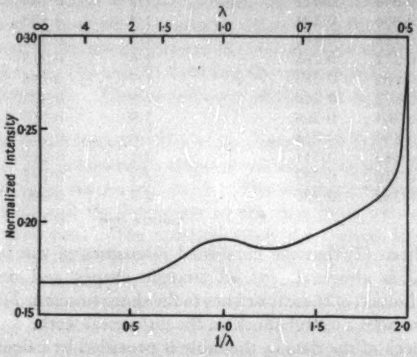


Figure 2. Normalized intensity of centre line of spectrum plotted as a function of λ. λ is the ratio of the length of the equal sides of the isosceles triangle to the length of the third side.

§ 3. INTERMOLECULAR BROADENING

The spectrum calculated in § 2 refers to isolated triangle groups. However, in any actual crystalline specimen containing triangular groups the dipolar interaction between neighbouring groups broadens the spectrum and smooths the discontinuities. Pake (1948) has shown that this intermolecular broadening may be taken in account by multiplying the intramolecular spectrum by a symmetrical bell-shaped curve representing the spectrum of the intermolecular interactions. Let this normalized, even, broadening function be $g(h)$. Then the broadened spectrum is described by a shape function $G(h)$ given by

$$G(h) = \int_{-\infty}^{\infty} F(h_0) g(h - h_0) dh_0. \qquad \ldots\ldots(23)$$

For any given broadening function the broadened function $G(h)$ may be obtained from (23) by a numerical integration at each value of h.

The second moment of the broadened spectrum is the sum of the second moments of the unbroadened spectrum and of the broadening function.

This conclusion, which has often been assumed without proof, may be derived in the following manner. The second moment of the broadened function is

$$\begin{aligned} S_2 &= \int_{-\infty}^{\infty} h^2 G(h)\,dh \\ &= \int_{-\infty}^{\infty} dh_0 F(h_0) \int_{-\infty}^{\infty} h^2 g(h-h_0)\,dh. \qquad \ldots\ldots(24) \end{aligned}$$

Now

$$\begin{aligned} \int_{-\infty}^{\infty} h^2 g(h-h_0)\,dh &= \int_{-\infty}^{\infty} (h+h_0)^2 g(h)\,dh \\ &= \int_{-\infty}^{\infty} h^2 g(h)\,dh + 2h_0 \int_{-\infty}^{\infty} h g(h)\,dh + h_0^2 \int_{-\infty}^{\infty} g(h)\,dh. \qquad \ldots(25) \end{aligned}$$

The first term on the right in (25) is the second moment S_{2b} of the intermolecular broadening function; the second term is zero on account of the even property of $g(h)$; the third term is h_0^2 since $g(h)$ is normalized. Consequently (24) becomes

$$\begin{aligned} S_2 &= \int_{-\infty}^{\infty} (h_0^2 + S_{2b}) F(h_0)\,dh_0 \\ &= \int_{-\infty}^{\infty} h_0^2 F(h_0)\,dh_0 + S_{2b} \\ &= S_{2a} + S_{2b}, \qquad \ldots\ldots(26) \end{aligned}$$

where S_{2a} is the second moment of the unbroadened, intramolecular spectrum; thus the theorem is proved.

It has been customary to assume that the broadening function $g(h)$ should be Gaussian, taking the form

$$g(h) = (2\pi\beta^2)^{-1/2} \exp(-h^2/2\beta^2). \qquad \ldots\ldots(27)$$

The second moment S_{2b} of this function is β^2. Van Vleck (1948) has pointed out that a Gaussian function can never be quite correct since its fourth moment is too large. The fourth moment of the Gaussian function is

$$S_{4b} = \int_{-\infty}^{\infty} h^4 g(h)\,dh = 3\beta^4. \qquad \ldots\ldots(28)$$

Thus the ratio the fourth moment to the square of the second moment is

$$S_{4b}/(S_{2b})^2 = 3,$$

for a Gaussian function. Van Vleck has shown that for a dipolar broadened spectrum this ratio must always be less than 3; for the only case which he worked out explicitly, namely a simple cubic lattice with the magnetic field applied along the (100) direction, the ratio was found to be 2·44. Thus although a Gaussian function is not greatly in error, a more correct function would have a flatter top and more strongly convergent sides. A broadening function using a fourth power in the exponent satisfies this requirement qualitatively. This function is

$$g(h) = [\tfrac{1}{2}\gamma\Gamma(\tfrac{1}{4})]^{-1} \exp(-h/\gamma)^4; \qquad \ldots\ldots(29)$$

the coefficient in front of the exponential is required for normalization. The fourth moment is $\gamma^4/4$ and the second moment is $\gamma^4\Gamma(\frac{3}{4})/\Gamma(\frac{1}{4})$; the ratio $S_{4b}/(S_{2b})^2$ for this function is thus found to be 2·19. It is likely therefore that the correct broadening function lies in most cases between those described by (27) and (29).

§ 4. Application to the Oxonium Ion

The proton magnetic resonance spectra of a number of acid hydrates have been measured by Richards and Smith (1951). They found that the spectra of the monohydrates of nitric, perchloric and sulphuric acids could all be closely accounted for by the broadened spectrum of an equilateral triangle of protons with a reasonable length of side, 1·72 Å. On the other hand none of the spectra could be fitted on the basis of pairs of water protons and remote acid protons. It was therefore concluded that these acids contained the oxonium ion (H_3O^+). Calculations were not then available for triangles distorted from equilateral form, and it was therefore not possible to conclude that the oxonium ion is equilateral since the data might equally well have been fitted by triangles of other shapes. In this section we shall attempt to fit the observed spectra to the theory for isosceles triangles in order to see how specific the agreement with equilateral-triangle theory is as a test of equilateral shape. If the oxonium ion in these hydrates is in fact not equilateral there is no special reason why it should be isosceles rather than of general shape. However the theory for a triangle of general shape, which would require an additional parameter to describe the triangle, would be more complicated than that for the isosceles triangle. Moreover it may be expected that if a spectrum can be fitted for an isosceles triangle of given λ it will also be possible to fit it for non-isosceles triangles having a comparable distortion from equilateral shape.

A measured spectrum of nitric acid monohydrate is shown in figure 3. (The details of this spectrum and for perchloric acid monohydrates shown in figure 4

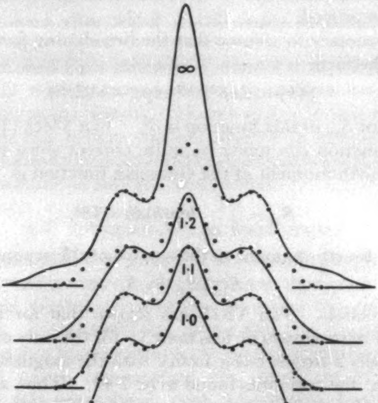


Figure 3. Comparison of the measured proton resonance spectrum for nitric acid monohydrate (Richards and Smith 1951) with calculated spectra (full lines) for an equilateral oxonium ion ($\lambda=1$), isosceles ions having $\lambda=1\cdot1$ and $1\cdot2$, and for water molecules with remote acid protons ($\lambda=\infty$). The value of λ is shown within the apex of each calculated spectrum. The four calculated curves are displaced from each other for clarity and the experimental points are shown with each curve.

were kindly provided by Dr. Richards.) Also shown in figure 3 are theoretical spectra for $\lambda = 1$, 1·1, 1·2 and ∞, broadened by a Gaussian function in the manner described in § 3. The parameters α and β have been chosen by trial to give the best agreement between theory and experiment. The curve for $\lambda = \infty$, corresponding to a pair of water protons and a remote acid proton, differs markedly from the experimental spectrum. Some reduction in the discrepancy at the central peak could be achieved by the use of a broader broadening function (larger β), but only at the expense of the agreement on the shoulders of the spectrum and of a still greater discrepancy in the wings. On the other hand the agreement for $\lambda = 1$, though not perfect, is good, especially on the shoulders and in the wings where diagreement would cause a large discrepancy in second moment. The agreement for $\lambda = 1{\cdot}1$ is also quite good, though rather less satisfactory than for $\lambda = 1$. On the other hand for $\lambda = 1{\cdot}2$ the agreement is not good, since there is an important discrepancy at the shoulders and in the wings. If the parameter α is made greater in order to improve agreement at the shoulders the discrepancy in the wings becomes worse. Any attempt at improvement by reducing β makes the central peak too tall. The curve shown represents the best compromise. We conclude from figure 3 that for nitric acid monohydrate λ cannot exceed 1·1 if reasonable agreement is to be obtained between experiment and theory, using a Gaussian broadening function.

A measured spectrum of perchloric acid monohydrate is shown in figure 4 together with calculated spectra for $\lambda = 1$, 1·1, and 1·2 using Gaussian broadening.

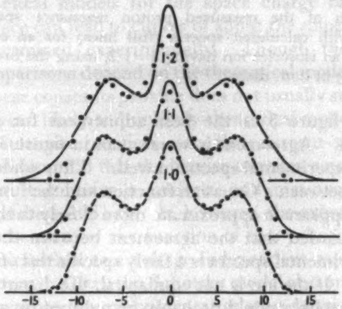


Figure 4. Comparison of the measured proton resonance spectrum for perchloric acid monohydrate (Richards and Smith 1951) with calculated spectra (full lines) for an equilateral oxonium ion ($\lambda = 1$), and for isosceles ions having $\lambda = 1{\cdot}1$ and 1·2. The value of λ is shown within the apex of each calculated spectrum. The three calculated curves are displaced from each other for clarity, and the experimental points are shown with each curve.

The spectrum for $\lambda = \infty$, which disagrees as grossly as in figure 3, is not shown. Here again agreement is good for $\lambda = 1$. The discrepancy for $\lambda = 1{\cdot}2$ is pronounced and is more evident for $\lambda = 1{\cdot}1$ than for nitric acid monohydrate. We can therefore say for perchloric acid monohydrate that the oxonium ion must be even more nearly equilateral in nitric acid monohydrate.

In § 3 it was seen that the use of a Gaussian broadening function is not quite correct, and that the proper function is likely to be intermediate in form between a Gaussian function and the function described by (29). The latter function has a flatter top and more convergent wings than the Gaussian function and this is qualitatively what is required to bring about agreement between the experimental spectra and the theoretical spectrum for $\lambda = 1{\cdot}2$. However, using this broadening function, the best adjustment of the theory for $\lambda = 1{\cdot}2$ to the experimental spectrum for nitric acid monohydrate is shown in figure 5. It will be seen that the agreement is little better than in figure 3. We may conclude therefore that no broadening function which approximates to the correct function will effect agreement between theory and experiment for $\lambda = 1{\cdot}2$ for nitric acid monohydrate. This result applies equally to perchloric acid monohydrate.

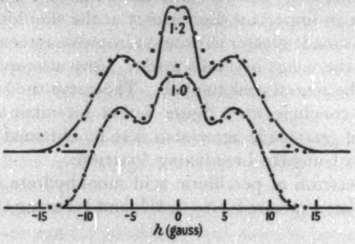


Figure 5. Comparison of the measured proton resonance spectrum for nitric acid monohydrate with calculated spectra (full lines) for an equilateral oxonium ion ($\lambda = 1$), and for an isosceles ion having $\lambda = 1{\cdot}2$, using the broadening function given by equation (29) in both cases.

Also shown in figure 5 is the best adjustment for $\lambda = 1$ using the new broadening function. Agreement is worse than in figure 3 where the Gaussian function fitted the experimental spectrum well. Thus while the true broadening function should lie between a Gaussian function and the function with the fourth power exponent, it appears to approximate more closely to the Gaussian function.

It is finally concluded that the agreement between the equilateral triangle theory and the experimental spectra is a fairly specific test of the equilateral shape of the oxonium ion. If the ion is not equilateral, the departure is less than 10%. This margin of uncertainty could probably be reduced by measuring the spectra of single crystals of the acid hydrates. The best values of α obtained in fitting the spectra to the theory for $\lambda = 1$ were 4·25 G for nitric acid monohydrate and 4·33 G for perchloric acid monohydrate, corresponding to equilateral triangles whose sides R are 1·71 Å and 1·70 Å respectively. With $\lambda = 1{\cdot}1$ the best agreement is obtained with triangles having $R = 1{\cdot}59$ Å and $P = 1{\cdot}75$ Å.

REFERENCES

ANDREW, E. R., and BERSOHN, R., 1950, *J. Chem. Phys.*, **18**, 159.
PAKE, G. E., 1948, *J. Chem. Phys.*, **16**, 327.
RICHARDS, R. E., and SMITH, J. A. S., 1951, *Trans. Faraday Soc.*, **47**, 1261.
VAN VLECK, J. H., 1948, *Phys. Rev.*, **74**, 1168.

Current–Capacitance Characteristics of Metal Rectifiers

By YOSHIRO MORIGUCHI† AND ATUYOSI OKAZAKI‡

† Kurume University ‡ Faculty of Science, Kyusyu University

Communicated by P. T. Landsberg ; MS. received 3rd May 1957, and in final form 29th May 1957

Abstract. Using a substitution bridge method, the reverse current–voltage and capacitance–voltage characteristics of metal rectifiers have been measured at room and liquid air temperatures. At low temperatures the capacitance of Cu_2O and Se rectifiers, and the reverse current of the former, decrease, but that of the latter increases. The capacitance–voltage characteristics seem to be in good agreement with Landsberg's heterogeneous barrier layer theory, and the comparison of the experimental results with this theory gives effective dielectric constants of 4·3 and 2·0∼4·3 for the barrier materials of Cu_2O and Se rectifiers, respectively.

§ 1. INTRODUCTION

THE correlation of experimental and theoretical current–voltage relations in metal rectifiers involves the following well-known difficulty. A variety of different theoretical models for the space charge barrier at the metal–semiconductor interface can lead to curves of about the same shape, and in agreement with that determined experimentally. Though the rectifier constants obtained by such comparisons depend on the theoretical model used, the numerical information which these constants provide does not usually suffice to make a really clear distinction between likely and unlikely theoretical models. This holds true for copper oxide and selenium rectifiers, since their current–voltage characteristics can be fitted on the assumption that the barrier contains a negligible space charge (Mott type), or that it contains a uniform and non-negligible space charge (Schottky type) (Landsberg 1951 a). If, however, the experimental curve giving the capacitance C of the rectifier as a function of bias voltage U is also used, then the Mott barrier, which leads to a voltage-independent capacitance, is normally ruled out on that account. Often the Schottky model also proves to be inadequate by this test, either because the experimental relation between C^{-2} and U is not linear, as would be expected on this model, or because the intercept of the experimental curve on the $C^{-2}=0$ axis is of the wrong magnitude. (It should be of the order of the diffusion potential.) Thus the test of theoretical model by both current and capacitance characteristic is far more searching. On the theoretical side, however, adjustable parameters, which specify the space charge distribution and are as yet incapable of being checked independently to a satisfactory extent, still remain in any theoretical treatment which is based on a special model for the barrier. In order to reduce the number of adjustable parameters, which rob the comparison of theory with experiment of much of its usefulness, it is therefore desirable to have a theory which, while using both the experimental current–voltage and capacitance relations, is independent of the precise space-charge distribution in the barrier.

A theory of this type, due to Landsberg (1951 b), predicts that for a large class of metal rectifiers, whatever their precise space charge barrier, a simple relation can be found between the capacitance–bias voltage curve and the current–bias voltage curve, which enables one to obtain the latter if the former is given. The main predictions were confirmed by Lees (1953 b) for cuprous oxide rectifiers. This work is extended here to a greater temperature range and to selenium rectifiers.

§ 2. THEORETICAL BASIS

According to Landsberg's theory (Landsberg 1951 a, b, 1952) of metal rectifiers with a single carrier and a general distribution of impurity centres in the barrier layer, the reverse current density j when the reverse voltage $U(<0)$ is applied is given by

$$\frac{j}{j_0}=y^{3/2}e^{y}\left[1-\exp\left(\frac{\boldsymbol{e}U}{\boldsymbol{k}T}\right)\right], \qquad \ldots\ldots(1)$$

where

$$j_0=\left(\frac{2}{\pi}\right)^{1/2}n(d)v\epsilon\left(\frac{\boldsymbol{k}T}{\boldsymbol{e}}\right)^2\exp\left(-\frac{\boldsymbol{e}V_{\mathrm{D}}}{\boldsymbol{k}T}\right), \qquad \ldots\ldots(2)$$

$$y\equiv y(U)=\boldsymbol{e}^2/2\epsilon b(U)\boldsymbol{k}T, \qquad \ldots\ldots(3)$$

and ϵ is the dielectric constant, $n(d)$ the concentration of the impurity centres in the bulk semiconductor, v the mobility of the current carriers, V_{D} the 'diffusion potential' and $b(U)$ the distance, measured into the semiconductor from the interface, of the plane at which the potential energy of the current carriers is a maximum.

In addition, the capacitance per unit area of the barrier layer is given by

$$C=\frac{1}{4\pi}\frac{(\epsilon\boldsymbol{k}T)^2}{\boldsymbol{e}^3}\frac{d(y^2)}{d(V_{\mathrm{D}}-U)}, \qquad \ldots\ldots(4)$$

and on integrating equation (4) we have

$$4\pi\frac{\boldsymbol{e}^3}{(\epsilon\boldsymbol{k}T)^2}\int_U^0 C\,dU=y^2(U)-y^2(0). \qquad \ldots\ldots(5)$$

For values of U smaller than about $-0{\cdot}1$ v, $\exp(\boldsymbol{e}U/\boldsymbol{k}T)\ll 1$, so that we can rewrite equations (1) and (5) as follows:

$$\tfrac{3}{2}\ln y(U)+y(U)=\ln j(U)-\ln j_0, \qquad \ldots\ldots(6)$$

and

$$\ln[y^2(U)-y^2(0)]=\ln\left[4\pi\frac{\boldsymbol{e}^3}{(\boldsymbol{k}T)^2}\int_U^0 C\,dU\right]-2\ln\epsilon. \qquad \ldots\ldots(7)$$

Here $\frac{3}{2}\ln y(U)+y(U)$ and $\ln[y^2(U)-y^2(0)]$ are connected through the parameter $y(U)$, and one can draw the theoretical curves of $\frac{3}{2}\ln y(U)+y(U)$ as function of $\ln[y^2(U)-y^2(0)]$. These depend on the parameter $y^2(0)$ as shown in figure 3. On the other hand, using the measured j and C values for each reverse voltage U, one can plot the experimental curve of $\ln j$ as function of

$$\ln\left[4\pi\frac{\boldsymbol{e}^3}{(\boldsymbol{k}T)^2}\int_U^0 C\,dU\right].$$

It is seen from equations (6) and (7) that the latter experimental curve should coincide with one of the theoretical curves by parallel displacements of $2\ln\epsilon$ horizontally and $\ln j_0$ vertically. From the values of these displacements one can find values of ϵ and j_0. This type of analysis was also used by Lees (1953 a).

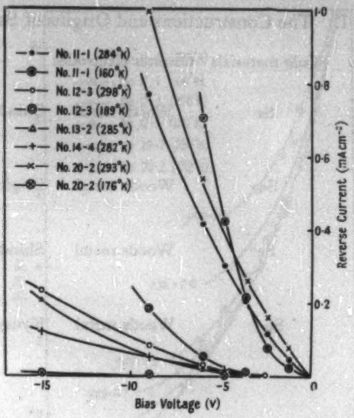


Figure 1. Reverse current–bias voltage characteristics.

In this paper, the reverse voltage–current and reverse voltage–capacitance characteristics of several kinds of Cu_2O and Se rectifiers have been studied at room and liquid air temperatures, in order to determine whether equations (6) and (7) hold good for these rectifiers.

§ 3. EXPERIMENTAL PROCEDURE AND RESULTS

The samples used for the measurements were several Cu_2O and Se rectifiers of different origins and constructions as listed in table 1. These samples were mounted in a desiccator which contained $CaCl_2$, and were kept *in vacuo* at low temperature in order to eliminate the influence of humidity.

A.C. characteristics were measured by a substitution bridge method using a small a.c. voltage (frequency 1000 c/s) superposed on a d.c. bias voltage. The series resistances of the rectifiers used were so small compared with the barrier impedances that the impedance locus diagram method was unnecessary to determine the capacitances of the barrier.

When a large reverse voltage was applied to the metal rectifiers, the 'creep' phenomena occurred so that the measurements of d.c. and a.c. characteristics were difficult. In order to avoid such complications, every sample was electrically 'formed' by the application of a sufficiently high reverse voltage. In spite of this treatment, some samples, especially at low temperature, showed a sudden decrease of the capacitance caused by creeping at the bias voltage of several volts. For these samples the capacitance at the higher bias voltage was obtained by extrapolation along the dotted lines of figure 2 in order to calculate the integral in equation (7). But this adjustment had only a very small effect in the case of all samples in which it was necessary, excepting some Se rectifiers at low temperature. The reverse voltage–current characteristics for some samples are shown in figure 1. It is found that the reverse current of Se rectifiers at low temperature is larger than that at room temperature, while the situation is reversed in the case of Cu_2O rectifiers. In figure 2, the voltage–capacitance characteristics

Table 1. The Constructions and Origins of Samples

Number of samples	Bulk materials	Counter electrodes	Origins
10–1 11–1 11–2	Se	Woods metal	Shindengen Kogyo K.K.
12–1 12–2 12–3	Se	Woods metal	Origin Denki K.K.
13–1 13–2	Se	Woods metal	Shindengen Kogyo K.K.
14–1 14–2 14–3 14–4	Se †	Woods metal	Kyusyu University
20–1 20–2 20–3 20–4 20–5	Cu_2O	Cu	Origin Denki K.K. and Shindengen Kogyo K.K.‡

† There were artificial layers of CdSe between the Se and the electrode, having thicknesses, respectively, of 0·48 μ (14–1), 0·86 μ (14–2), 1·0 μ (14–3), and 2·2 μ (14–4).

‡ There was no appreciable difference in d.c. and a.c. characteristics between Cu_2O samples from Origin Denki K.K. and Shindengen Kogyo K.K.

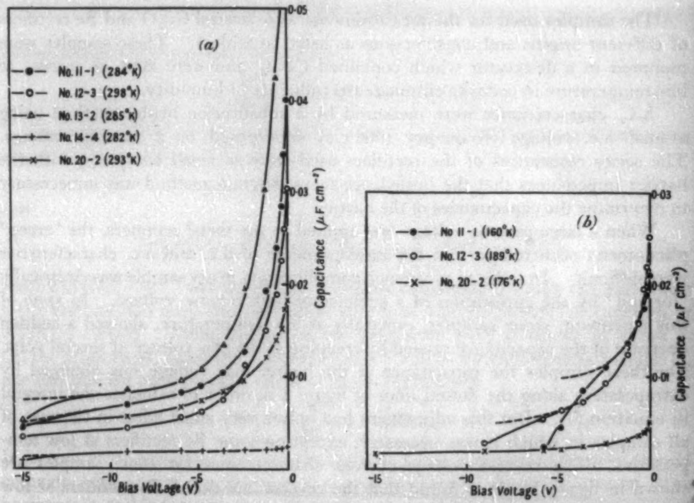


Figure 2. (*a*) Capacitance at room temperature. (*b*) Capacitance at low temperature.

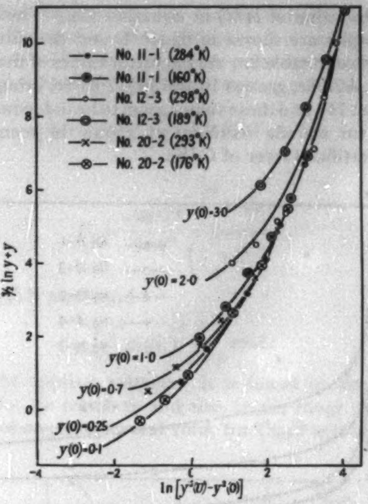


Figure 3. ($\frac{3}{2}\ln y+y$, $\ln[y^2(U)—y^2(0)]$) curves and measured points of $\ln j$, $\ln[4\pi e^3/(kT)^2\int_U^0 CdU]$.

for some samples are shown. Two facts deserve notice. The first is that both the capacitance itself, and its voltage dependence, are smaller at low temperatures and the second is that the sudden decrease of the capacitance of the Se rectifiers caused by creep occurs at smaller bias voltage at the lower temperatures. The former fact has already been reported by Niimi (1955) and Moriguchi (1955 a), and the latter seems to be related to the decrease of the resistance at low temperatures, but we can give no reasonable explanation for it (Pogansky 1949, Hoffmann 1950, Oldekop 1952, Moriguchi 1955 b). The voltage–capacitance relation of Se rectifiers of the group 14 which has an artificial layer of CdSe between Se and counter-electrode is quite different from those of others. This point, however, will be discussed in detail elsewhere.

§ 4. ANALYSIS OF RESULTS

In figure 3 the theoretical values of $\frac{3}{2}\ln y+y$ are plotted against the values of $\ln[y^2(U)-y^2(0)]$, using equations (6) and (7) with parameters $y(U)$ and $y(0)$. Measured values of $\ln j$ are also shown as a function of

$$\ln\left[4\pi\frac{e^3}{(kT)^2}\int_U^0 CdU\right]$$

after suitable parallel displacements to attain the best possible overlap of each experimental curve with one of the theoretical curves. As seen in the figure, the measured points all fall on the theoretical curve very well. From this figure, one can determine the value of $y(0)$ as well as the values of ϵ and j_0 as mentioned in § 2. Hence one can find the value of $y(U)$ in equation (7), and, using the value of

ϵ, one can obtain the value of $b(U)$ in equation (3). The ($b(U)$, bias voltage) curves for some samples are shown in figure 4, and the values of ϵ, j_0, and $b(0)$ for all samples are listed in table 2. As the mean values of the dielectric constants ϵ, we get 4·3 (Cu_2O), 4·3 (Se, groups 10 and 11), 2·0 (Se, group 12), 2·9 (Se, group 13), and 3·3 (Se, group 14), and these values seem to be independent of temperature. The value of $b(0)$ for the Se rectifiers of group 14 seems to increase with the thickness of the artificial layer of CdSe.

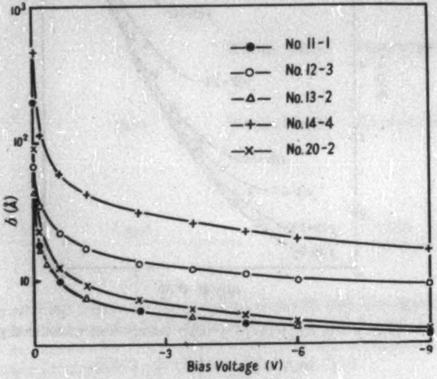


Figure 4. Inferred position of the potential maximum in the (barrier, bias voltage) curves at room temperature.

Table 2. The Values of ϵ, j_0, and $b(0)$

Samples	Temperatures (°K)	ϵ	j_0 (μA cm^{-2})	$b(0)$ (Å)
10–1	284	4·4	0·40	167
11–1	284	4·4	5·2	200
11–1	160	4·5	1·8	116
11–2	282	4·0	0·93	93
12–1	300	2·0	$1{\cdot}0\times10^{-3}$	70
12–2	278	2·1	$2{\cdot}5\times10^{-3}$	55
12–3	278	2·0	$1{\cdot}6\times10^{-3}$	70
12–3	189	2·0	$6{\cdot}7\times10^{-5}$	74
13–1	285	2·9	$8{\cdot}2\times10^{-4}$	39
13–2	285	2·9	$8{\cdot}2\times10^{-4}$	44
14–1	282	3·7	$1{\cdot}0\times10^{-2}$	160
14–2	282	3·4	$1{\cdot}3\times10^{-2}$	291
14–3	282	3·4	$1{\cdot}3\times10^{-2}$	291
14–4	282	2·5	$1{\cdot}8\times10^{-2}$	476
20–1	284	3·8	16	154
20–2	293	4·5	32	95
20–2	176	4·6	$8{\cdot}7\times10^{-3}$	41
20–3	293	4·5	68	64
20–4	293	4·5	68	105
20–5	293	3·9	16	122

Let n and $V(x)$ be, respectively, the density of the ionized impurity centres and the potential in the barrier whose thickness is d, and the plane $x=0$ be the interface. If a class A or Schottky barrier is presumed, according to Landsberg (1951 b), we have the following capacitance–voltage relation:

$$\frac{1}{C^2}=(2\pi)^2\left(\frac{2e}{\epsilon kT}\right)^4 eV_0(V_{\mathrm{D}}-U), \qquad \ldots\ldots(8)$$

where

$$V_0=\frac{\epsilon^3k^4T^4}{4\pi e^6}\frac{\int_0^d xn\,dx}{\left(\int_b^d n\,dx\right)^2}, \qquad \ldots\ldots(9)$$

where $V'(d)$ is negligibly small and

$$\left(\int_b^d n\,dx\right)^2\Big/\int_0^d xn\,dx$$

is independent of the applied voltage. It is found in figure 5 that the above relation is satisfied for Se rectifiers and also, in the range of the applied voltage from several tenths of a volt to several volts, for Cu_2O rectifiers.

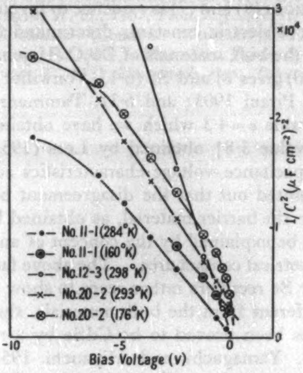


Figure 5. $1/C^2$ against bias voltage relations. (The ordinate for No. 20–2 (176°K) should be read ×10.)

§ 5. Discussion

The reverse current density of Se rectifiers increases at low temperatures, as seen in figure 1. This effect is well known, but difficult to explain theoretically. In trying to explain this effect, one must make at least one of the following assumptions: (1) that the diffusion potential V_{D} decreases, or (2) that the carrier density $n(d)$ increases, or (3) that the carrier mobility v increases at low temperatures. But the diffusion potential V_{D} seems to increase at low temperatures (Cooper 1952, Matukura and Okazaki 1954, Moriguchi 1955 a), and the product $en(d)v$, namely the bulk conductivity of Se, seems to decrease at low temperatures

(Plessner 1951 a, b, Henkels and Maczuk 1953). But it is possible that if, for the temperature range in question, all impurities are ionized in the barrier, so that $n(d)$ in equation (2) is temperature independent, the effect can to some extent be explained as follows (Landsberg 1952, § 3). The intersection of characteristics belonging to different temperatures may be due to a minimum through which the mobility of micro-crystalline selenium passes as the temperature is lowered. Plessner found indications of such a minimum.

The capacitance of the barrier layer is temperature dependent in both Cu_2O and Se rectifiers, as seen by comparing figures 2(*a*) and 2(*b*), in contrast with Lees's findings of a temperature independent capacitance for Cu_2O rectifiers. But he covered a narrower temperature range. This temperature dependence cannot be explained in terms of a temperature dependence of the diffusion potential. Assuming the same charge distribution at all temperatures and all bias voltages, we would have to have V_D two to four times greater at liquid air temperature than it is at room temperature. But the value of j_0 would then change by more than a factor of 10, owing to the term $\exp(-\boldsymbol{e}V_D/\boldsymbol{k}T)$ in equation (2), and this is not the case as seen from table 2. A more likely explanation is that under differential expansion and reorientation of microcrystallites the impurity distribution in the barrier changes. This would give rise to a temperature dependent capacitance.

The good agreement of the measured characteristics with the theoretical curves given by equations (6) and (7) is evidence in favour of Landsberg's theory. But the values of the dielectric constants determined from this agreement are smaller than those of the bulk materials of Cu_2O (Højendahl (1933) finds a value of 10·5; Wright (1950) gives 8) and Se (6·13, Vonwiller and Mason 1907; 6·60, Schmidt 1903; 7·44, Pirani 1903; and 6·31, Tammann and Boehme 1931, for vitreous Se). The value $\epsilon = 4·3$ which we have obtained for Cu_2O is in good agreement with the value 3·8† obtained by Lees (1953 b), who also analysed his results for the capacitance–voltage characteristics according to Landsberg's theory. Lees has pointed out that the disagreement between the value of the dielectric constant for the barrier material, as obtained by him, and that for the bulk material cannot be explained by the concept of an 'effective' contact area smaller than the geometrical contact area. The above fact and the various values of ϵ obtained here for Se rectifiers rather seem to show that the barrier material may be somewhat different from the bulk material; the barrier material of the usual Se rectifiers has been proved to be CdSe by x-ray analysis (Yamaguchi and Katayama 1950, Yamaguchi and Miyauchi 1951), and in preliminary measurements the present authors obtained a value of about 3 for the dielectric constant of CdSe.

§ 6. CONCLUSION

From the d.c. and a.c. measurements on Cu_2O and Se rectifiers, we may conclude as follows: (1) Landsberg's relation between the reverse current and the capacitance is satisfied very well both in Cu_2O and Se rectifiers. (2) The following points remain unexplained, and require further study: (*a*) the dielectric constants determined from the current–capacitance characteristic are smaller than those of bulk materials, and (*b*) the temperature dependence of the capacitance is not sufficiently explained.

† In an improved analysis Lees (1953 a) adopted the value 4·3, which is identical with ours.

ACKNOWLEDGMENTS

We wish to express our thanks to Dr. P. T. Landsberg for his valuable comments. We are also indebted to Shindengen Kogyo K. K., Origin Denki K. K., Professor H. Tubota and Mr. H. Suzuki for the samples used.

REFERENCES

COOPER, R., 1952, *Proc. Phys. Soc.* B, **65**, 409.
HENKELS, H. W., and MACZUK, J., 1953, *Phys. Rev.*, **91**, 1562.
HOFFMANN, A., 1950, *Z. angew. Phys.*, **2**, 363.
HØJENDAHL, K., 1933, *Z. phys. Chem.* B, **20**, 54.
LANDSBERG, P. T., 1951 a, *Proc. Roy. Soc.* A, **206**, 463; 1951 b, *Ibid.*, **206**, 477; 1952, *Proc. Phys. Soc.* B, **65**, 397.
LEES, J., 1953 a, Ph.D. *Thesis*, University of Aberdeen; 1953 b, *Proc. Phys. Soc.* B, **66**, 622.
MATUKURA, Y., and OKAZAKI, A., 1954, *Mem. Fac. Sci., Kyusyu University* B, **1**, 4, 119.
MORIGUCHI, Y., 1955 a, *Kurume University J.*, **6**, 10; 1955 b *Ibid.*, **6**, 17 (in Japanese).
NIIMI, T., 1955, *J. Phys. Soc., Japan*, **10**, 445.
OLDEKOP, W., 1952, *Z. Phys.*, **134**, 66.
VON PIRANI, M., 1903, *Dissertation*, Humboldt-Universität, Berlin.
PLESSNER, K. W., 1951 a, *Proc. Phys. Soc.* B, **64**, 671; 1951 b, *Ibid.*, **64**, 681.
POGANSKY, S., 1949, *Dissertation*, Technische Universität, Berlin–Charlottenburg.
SCHMIDT, W., 1903, *Ann. Phys., Lpz.*, **4**, 11, 119.
TAMMANN, G., and BOEHME, W., 1931, *Z. anorg. Chem.*, **197**, 1.
VONWILLER, O. U., and MASON, W. H., 1907, *Proc. Roy. Soc.* A, **79**, 177.
WRIGHT, D. A., 1950, *Semiconductors* (London: Methuen) p. 42.
YAMAGUCHI, J., and KATAYAMA, S., 1950, *J. Phys. Soc., Japan*, **5**, 387.
YAMAGUCHI, J., and MIYAUCHI, T., 1951, *Electrotech, Pap., Tokyo*, **3**, 5.

RESEARCH NOTES

A New Colloid for use with the Bitter Figure Technique

By D. J. CRAIK† and P. M. GRIFFITHS
Department of Physics, University of Nottingham

MS. received 27th June 1957

During recent studies of the domain configurations of ferromagnetic materials by the Bitter figure technique (1931), difficulties were experienced due to heavy staining of the specially prepared specimen surface. The technique entails the spreading of a thin film of a colloidal suspension of magnetite, Fe_3O_4, between the surface and a thin glass cover slip. The colloidal particles are attracted to any magnetic poles which appear at the surface due to the intersection of a domain wall with the surface, and to scratches or non-ferromagnetic inclusions in the surface. It is frequently desirable to observe the powder patterns for a prolonged period, but the time of observation is severely limited by the effects of the colloid on the surfaces. Two colloids are in general use, described by Elmore (1938) and by Bozorth (1951), respectively.

Elmore's colloid is prepared by dissolving 2 g $FeCl_2.4H_2O$ and 5·4 g $FeCl_3.6H_2O$ in 300 cm^3 of water at 70°c, to which with constant stirring is added 5 g NaOH dissolved in 50 cm^3 of water. A heavy precipitate of magnetite slurry is formed, and this is filtered and washed with distilled water to remove salt and excess NaOH ; it is then peptized by washing with 0·01N solution of HCl. The precipitate is next added to a 0·5% solution of sodium oleate and boiled until the solution is colloidal.

Bozorth prepares the magnetite slurry in the same manner as above, but instead of peptizing the slurry, it is added to a concentrated solution of dodecylamine. The dodecylamine solution is prepared by adding 2 g to about 10 cm^3 of normal HC1 solution until a pH of 7, as indicated by pH papers, is reached. This solution is then diluted with distilled water to 50 cm^3, and 20 cm^3 of slurry is then added. The whole system is made up to 150 cm^3 with distilled water and stirred vigorously at about 6000 r.p.m. until the suspension becomes colloidal. It is finally diluted to 600 cm^3.

The main difference between the two colloids apparently lies in the particle size. In the case of Elmore's colloid the particles are well dispersed by peptization by the HCl, while Bozorth's colloid consists of a suspension of relatively large aggregates of the crystallites. These aggregates have been found by Heidenreich to be 1000 Å in diameter, and to be positively charged with respect to the solution. The individual crystallites have recently been measured using the electron microscope method by Craik (1956) to be of the order of 100 Å.

The staining of the metal specimen may be caused by two different mechanisms. First, any free Cl^- radicals present in the Elmore colloid will have a slight corrosive effect on the specimen surface, which will in time become appreciable. An attempt was made to remove Cl^- radicals from the colloidal solution, but this resulted in flocculation of the colloidal particles. With both colloids, the magnetite itself causes staining by being strongly adsorbed on the surface ; in like manner the magnetite will stain even a glass slide. It was noticed in some studies of wall movement in which a magnetic field was applied to the specimen, that after a time the colloid lost its apparent mobility, so that when wall movement occurred the whole of the colloidal deposit did not move with the wall but left traces where the wall had remained for a little time, thus giving an apparent doubling of its record.

---

† Now with Boots Pure Drug Co. Ltd., Beeston, Notts.

Figure 1 (Plate) shows a central 180° domain wall terminating at the edge of the photograph in a typical 90° wall triangular closure structure, and at the left-hand side is a second 180° domain wall. A magnetic field has been applied to the specimen and has driven both walls to the right. The new positions of the walls are given by heavy colloidal deposits, and their former positions are marked by fine colloidal traces. a, a′, b, b′ and c, c′ mark the positions of the two 180° domain walls and the 90° domain closure wall respectively ; a, b and c being their positions before a magnetic field has been applied and a′, b′ and c′ those in the presence of an applied magnetic field. Figure 2 (Plate) shows powder patterns on a (100) face of a silicon–iron specimen, where the heavy irregular stains around an inclusion are seen to mask the general domain structure.

When staining occurs, it can only be satisfactorily removed by regrinding the specimen surface. This entails rubbing lightly with grade 0000 emery, a process which leaves an amorphous surface layer which has to be removed by electropolishing and subsequent annealing. In an attempt to avoid this lengthy work we have examined a new colloid based on that used by Craik in his electron microscope study of powder patterns.

The slurry for the new colloid is prepared in the same manner as in the above methods. After careful washing with distilled water, it is added to a solution of 0·1N HCl. The slurry becomes peptized and a dense colloid is formed. It has been found that the best results are obtained if the slurry is used immediately after its preparation. An attempt made to peptize slurry which was several days old was unsuccessful. An equal amount of 1% solution of Celacol, a soluble derivative of cellulose, manufactured by the British Celanese Ltd., is then added. A small amount of glycerine, approximately 10% weight for weight of Celacol, is also added. The colloid is then dialysed against distilled water through a Cellophane membrane, until the Cl^- radicals are completely removed. This is tested by the addition of small amounts of silver nitrate solution to the dialysing water. The new colloid does not flocculate.

When this colloid was used for the prolonged examination (up to half an hour) of cobalt, and then removed by washing lightly in soapy water, microscope examination showed the surface to be stain-free. Even if the colloid was accidentally allowed to dry on a surface, provided the material was not readily oxidized, the dried film could be readily removed by washing, as before. The reason for the addition of the glycerine is that it reduces the adhesion of the dried film to the specimen surface.

Although it is clear that the removal of the acid prevents corrosion, it is not readily apparent why the magnetite itself does not now cause staining. It is considered that this is due to the complete protection which is afforded to the magnetite particles by the Celacol ; the same mechanism which prevents their mutual attraction and flocculation prevents their adhesion to the surface.

From the point of view of dispersion, the colloid is similar to that of Elmore, and is thus suitable for the observation of very fine structure. It does not, however, produce such an intense overall pattern as that of Bozorth. This is due to the fact that Brownian motion tending to a random distribution of the particles is more effective for very small particles, while attraction due to the magnetic inhomogeneities is greater for the larger aggregates in the Bozorth colloid. However, it has been found that the new colloid may be prepared in the same manner as recommended by Bozorth, i.e. the slurry may be added to the solution of Celacol and glycerine, and the slurry particles broken down by stirring. It may thus be used in its peptized state for the examination of fine structure detail, and in its stirred, larger aggregate state for the examination of the overall structure.

We wish to thank Professor L. F. Bates for his kind interest in this work. D. J. C. also wishes to thank Mr. W. J. Randall, of Boots Pure Drug Co. Ltd., for his continued interest and encouragement. P. M. G. wishes to thank the Department of Scientific and Industrial Research for a maintenance grant which has enabled him to carry out research in the subject to date.

References

Bitter, F., 1931, *Phys. Rev.*, **38**, 1903.
Bozorth, R. M., 1951, *Ferromagnetism* (New York : Van Nostrand), p. 533.
Craik, D. J., 1956, *Proc. Phys. Soc.* B, **69**, 647.
Elmore, W. C., 1938, *Phys. Rev.*, **54**, 309.

The Numerical Evaluation of the Frequency Response of Optical Systems

By H. H. HOPKINS
Department of Physics, Imperial College, London

MS. received 14*th June* 1957

§ 1. Introduction

In many problems of diffraction there occur double integrals of exponential functions of which the arguments may cause more or less rapid oscillations of the integrand over certain parts of the region of integration, which latter may also not be of simple analytical form. Both of these factors frequently give rise to difficulties in the evaluation of such integrals. An example of this kind of double integral arises in the calculation of the frequency response of an optical system with incoherent light from a knowledge of the pupil function of the system. For the reduced spatial frequency s, this requires the evaluation of the integral (Hopkins 1957),

$$D(s)=\frac{1}{A}\iint_{S}\exp\{iksW(x,y;\ s)\}dx\,dy \qquad \ldots\ldots(1)$$

where $k=2\pi/\lambda$, A is the area of the pupil, $W(x, y;\ s)$ is a polynomial in x and y, containing as many as ten terms, some of which may be of the fifth, or higher, degree in the variables, S is the area common to two pupil areas centred on the points $(\pm s/2, 0)$. For the axial image of a system having a circular pupil, the region S is that common to two circles having centres at these points. For off-axis images, particularly in the presence of vignetting, the region S is not of simple form, and has to be approximated, for example by a number of intersecting elliptical arcs. A further complication arises when the object is only partially coherent, for then the region S is further limited by the form of an 'effective source' (Hopkins 1953).

Integration based on Simpson's rule has been used for the evaluation of (1) in the case of primary spherical aberration (Black and Linfoot 1957). This method has the disadvantage of requiring very close intervals in regions where the argument $W(x, y;\ s)$ is increasing or decreasing rapidly, and it is precisely such regions in which the nett contributions to the value of the integral will be smallest. An alternative method, which is described below, is both elementary and easy to apply. It has already been used successfully in a number of problems, and is suitable for use in connection with digital computors.

§ 2. FORMULAE AND EXAMPLES

When the argument of the exponential function in the integral

$$I=\iint_S \exp\{if(x, y)\}dx\,dy \qquad \ldots\ldots(2)$$

is increasing or decreasing sufficiently rapidly, the precise form of its variation with x and y is of little consequence. Over a small enough region of the (x, y) plane, therefore, it will be the first two terms in the Taylor expansion of $f(x, y)$ that determine the contribution of that region to the value of the integral. Let the region be a rectangle defined by the lines

$$x=x_p\pm\epsilon_x, \qquad y=y_q\pm\epsilon_y.$$

Then, to the order of approximation contemplated, the mean value of the integrand of (2) over this rectangle will be given by

$$\frac{1}{4\epsilon_x\epsilon_y}\exp\{if(x_p, y_q)\}\int_{x_p-\epsilon_x}^{x_p+\epsilon_x}\exp\{if_x(x_p, y_q)(x-x_p)\}dx$$
$$\times\int_{y_q-\epsilon_y}^{y_q+\epsilon_y}\exp\{if_y(x_p, y_q)(y-y_q)\}dy$$

that is, by

$$\exp\{if(x_p, y_q)\}\frac{\sin\{\epsilon_x f_x(x_p, y_q)\}}{\{\epsilon_x f_x(x_p, y_q)\}}\frac{\sin\{\epsilon_y f_y(x_p, y_q)\}}{\{\epsilon_y f_y(x_p, y_q)\}}.$$

If now the contribution of any rectangle is accepted or rejected according as its centre (x_p, y_q) lies within or outside the region of integration, and the mean value of the integrand over the region bounded by the resulting zigzag is assumed to be equal to that over the exact region, the value of the integral (2) will be given by

$$I=(4\epsilon_x\epsilon_y)\sum_p\sum_q\exp\{if(x_p, y_q)\}\frac{\sin\{\epsilon_x f_x(x_p, y_q)\}}{\{\epsilon_x f_x(x_p, y_q)\}}\frac{\sin\{\epsilon_y f_y(x_p, y_q)\}}{\{\epsilon_y f_y(x_p, y_q)\}}. \qquad \ldots\ldots(3)$$

$4\epsilon_x\epsilon_y$ being the area of each elementary rectangle. If, in place of (2), the mean value of the integrand over S is required the factor $4\epsilon_x\epsilon_y$ is replaced by I/N, where N is the number of rectangles included in the double summation.

The formula (3) represents the contribution of each rectangle as given by the value of the integrand at its centre weighted by two factors, which are small, and possibly negative, when the first derivatives $f_x(x_p, y_q)$, $f_y(x_p, y_q)$ are large. Provided the integral exists, the right hand side of (3) always converges to the correct value, since in the limit $\epsilon_x \rightarrow 0$, $\epsilon_y \rightarrow 0$ it becomes simply the definition of the integral as the limit of a sum. Formulae similar to (3) could clearly be used for cases in which there are more than two variables of integration, provided the number of terms in the summations does not become prohibitive. For single integrals it is convenient to use as interval an exact sub-multiple of the region of integration.

In the case of pure defect of focus the argument of (1) has the simple form

$$W(x, y;\ s) = 2w_{20}x \qquad \ldots\ldots(4)$$

where w_{20} represents the path difference introduced for the marginal ray. The pupil is taken to be the interior of the circle $x^2+y^2=1$. An analytical expression for the corresponding integral has been obtained for this case (Hopkins 1955),

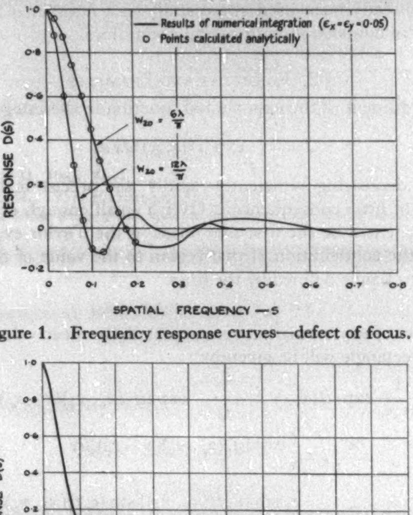


Figure 1. Frequency response curves—defect of focus.

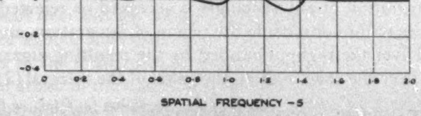


Figure 2. Frequency response curve for system with spherical aberration ($w_{60} = -11\lambda$, $w_{40} = +19\lambda$, $w_{20} = -9\lambda$).

but the resulting series of Bessel functions converges too slowly for practical computation for values of s much greater than $0{\cdot}5\lambda/w_{20}$, when $|w_{20}| > \lambda$. With the argument of the exponential function of the form (4), the first two terms of the Taylor expansion are, of course, an exact representation, so that the only error involved is that deriving from the approximate region of integration employed. It has nevertheless proved useful to carry through the details of the summations (3) even in this case. The results are shown in figure 1, together with points computed using the earlier series obtained analytically. In these, and the following computations, an interval $\epsilon_x = \epsilon_y = 0{\cdot}05$ has been used. A study of the influence of the size of interval is being carried out,† and results obtained so far suggest that this is a suitable size of interval for studies of frequency response where errors less than 0·01 are not of great practical importance.

The curve in figure 2 shows the frequency response for a case of primary and secondary spherical aberration combined with a defect of focus.

The form of the argument in (1) is then given by

$$W(x, y;\ s) = 2w_{20}x + w_{40}\{4(x^2+y^2)x + s^2x\} + w_{60}\{6(x^2+y^2)^2x + s^2(5x^2+3y^2)x + \tfrac{3}{8}s^4x\}$$

the chosen values of the aberration coefficients being $w_{20} = -11\lambda$, $w_{40} = +19\lambda$, $w_{60} = -9\lambda$. The resulting curve was obtained by a desk machine calculation. It is of interest to note that a consideration of aberration tolerances based on frequency response (Hopkins 1957) shows that $D(s) \geqslant 0{\cdot}75$ for $s = 0{\cdot}06$, and this is confirmed.

ACKNOWLEDGMENT

I am grateful to Mrs. K. Greenshields who carried out the computations.

REFERENCES

BLACK, G., and LINFOOT, E. H., 1957, *Proc. Roy. Soc.* A, **239**, 522.

HOPKINS, H. H., 1953, *Proc. Roy. Soc.* A, **217**, 408; 1955, *Ibid.*, **231**, 91; 1957, *Proc. Phys. Soc.* B, **70**, 449.

† By A. M. Goodbody of this department.

The Thermal Decomposition of Cupric Oxide *in vacuo*

BY A. GOSWAMI AND Y. N. TREHAN
National Chemical Laboratory, Poona 8, India

MS. received 18*th June* 1957, *and in final form* 12*th July* 1957

As is well known, the oxides of several metals decompose to a lower state or even to the metal when heated sufficiently. For example, Fe_2O_3 dissociates in air to Fe_3O_4 or FeO, at about 1350°c, and the decomposition of CuO to Cu_2O can be detected at 650°c, the reaction being complete at about 1100°c (Evans 1923, Mellor 1937, Ephraim 1948). Gulbransen and McMillan (1952) heated thin oxide layers supported on the parent metal *in vacuo* and attributed the change of CuO to Cu_2O to the presence of the metal. From thermodynamical considerations they concluded that the decomposition of CuO to Cu_2O *in vacuo* was not possible below 400°c.

We have found that oxidation of a copper surface at 350°c in air or at 100 or more mm Hg resulted in the formation of a layer of CuO over the initially formed Cu_2O on the metal surface. On the other hand, when the copper surface, after oxidation in air at 100 mm Hg pressure, was cooled *in vacuo*, the oxide layer consisted solely of Cu_2O, a result which suggests that dissociation of CuO had taken place during cooling.

This view was tested by heating cupric oxide *in vacuo*. Cupric oxide films, suitable for electron diffraction examination, were prepared by flaking off oxide coatings formed on copper and washing these in a 20% KCN solution followed by distilled water to remove traces of cuprous oxide and metallic copper. The typical transmission pattern (figure 1)†, indexed in table 1, is characteristic of pure CuO. After heating to 350°c *in vacuo* ($\sim 10^{-2}$ mm Hg or less) and subsequent cooling, also *in vacuo*, the flakes had changed from black to reddish brown in colour. The pattern (figure 2) reveals that they consisted of Cu_2O (table 2) with only traces of CuO.

Table 1. Analysis of the CuO pattern (figure 1)

Intensity	d_{obs} (Å)	hkl	d_{hkl} (Å)‡
vvf	2·834	011	2·836†
f	2·750	110	2·751
s	2·531	002	2·530
		$\bar{1}11$	2·523
s	2·318	111	2·323
		200	2·312
f	1·963	$\bar{1}12$	1·959
ms	1·858	$\bar{2}02$	1·866
vf	1·772	112	1·778
vvf	1·710	020	1·714
m	1·597	021	1·619
		202	1·581
m	1·508	$\bar{1}13$	1·505
m	1·421	022	1·418
f	1·393	$\bar{3}11$,310	1·410
		220,113	1·375
mf	1·312	311,$\bar{3}12$,221	1·304
mf	1·270	004	1·265
		$\bar{2}22$	1·261
vf	1·198	204,$\bar{1}14$	1·196
ms	1·170	$\bar{3}13$,222	1·167
vf	1·122	$\bar{4}02$,$\bar{2}23$	1·123
vf	1·108	$\bar{1}31$	1·092
f	1·019	024,223	1·018
f	0·9858	402	0·9921
		$\bar{2}24$,$\bar{1}15$	0·9808
vf	0·9430	$\bar{1}33$	0·9435
vf	0·9219	115,$\bar{3}31$	0·9209

s—strong, m—medium, f—faint, v—very.

† Calculated.

‡ Swanson and Tatge (1953).

In another experiment thin films of copper were condensed on to glass in a high vacuum ($\sim 10^{-5}$ mm Hg) and completely oxidized to CuO by heating in air. Subsequent heating and cooling *in vacuo* converted these films to Cu_2O.

These results show that CuO dissociates *in vacuo* at 350°c to Cu_2O even in the absence of metallic copper. It may thus be concluded that, contrary to the

---

† All figures are reproduced as Plates, at end of issue.

Table 2. Analysis of the Cu_2O pattern figure 2)

Intensity	d_{obs} (Å)	Cu_2O	
		hkl	d_{hkl} (Å)‡
m	3·045	110	3·020
vvf	2·746†	—	—
s	2·463	111	2·465
vvf	2·283†	—	—
m	2·143	200	2·135
vf	1·865†	—	—
f	1·739	211	1·743
vvf	1·625†	—	—
m	1·510	220	1·510
f	1·405	221,300	1·423
vvf	1·369†	—	—
m	1·292	311	1·287
f	1·238	222	1·233
vvf	1·162†	—	—
vvf	1·072†	—	—
vvf	1·035†	—	—
f	0·9809	331	0·9795
f	0·9554	420	0·9548
vvf	0·9230†	—	—
f	0·8742	422	0·8715
f	0·8250	511,333	0·8216

s—strong, m—medium, f—faint, v—very.
† Cupric oxide rings.
‡ Swanson and Fuyat (1953).

previously held view, the presence of metallic copper is not essential for this change to take place *in vacuo*.

It is of interest to mention that similar decomposition was observed even in the orientated CuO films. Thus, one-degree {110} orientated CuO crystals (figure 3) formed thermally at about 500°c and one-degree {021} orientated films (figure 4) formed anodically (Goswami 1954, 1956) possibly accompanied by {010} orientated crystals gave rise respectively to {110} (figure 5) and {100} (figure 6) orientated Cu_2O crystals on vacuum treatment. The change of orientation may be explained from their structures. In similar experiments, the decomposition of orientated α-Fe_2O_3 to orientated Fe_3O_4 and of Co_3O_4 to CoO has also been observed in this Laboratory (Goswami, unpublished work).

Recently Dixit and Agashe (1956) concluded that the oxide layer formed on a copper surface at 410°c and 75 mm Hg pressure consisted entirely of cuprous oxide. We have, however, observed that cupric oxide is formed even at 350°c and 75 mm Hg. It thus appears that the oxide formed at 410°c and 75 mm Hg was in fact cupric oxide, which decomposed to cuprous oxide during cooling *in vacuo*.

We are thankful to Professor G. I. Finch for his keen interest and advice during the course of this investigation.

References

Dixit, K. R., and Agashe, V. V., 1956, *Z. Naturf.*, **11a**, 41.
Ephraim, F., 1948, *Inorganic Chemistry* (London : Gurney and Jackson).
Evans, U. R., 1923, *Metals and Metallic Compounds* (London : Arnold).

Goswami, A., 1954, *J. Sci. Industr. Res.*, **13B**, 581 ; 1956, *Ibid.*, **15B**, 322.

Gulbransen, E. A., and McMillan, W. R., 1952, *J. Electrochem. Soc.*, **99**, 393.

Mellor, J. W., 1937, *A Comprehensive Treatise on Inorganic and Physical Chemistry* (London : Longmans Green) Vols III and XIII.

Swanson, H. E., and Fuyat, R. K., 1953, *Nat. Bur. Std. Circular* No. 539, **II**, 23.

Swanson, H. E., and Tatge, E., 1953, *Nat. Bur. Std. Circular* No. 539, **I**, 50.

Magnetoresistance in Strong Magnetic Fields

By R. BARRIE

Services Electronics Research Laboratory, Baldock, Herts.

MS. received 24*th July* 1957

The longitudinal magnetoresistance of solids in strong magnetic fields has been calculated by Argyres and Adams (1956) for the case in which the field H is so strong that $\hbar\omega_0 \gg kT$ and $\omega_0\tau \gg 1$ where $\omega_0 = eH/mc$ and τ is the time of relaxation for the scattering process. It is the purpose of this note to give the solution with the first restriction removed.

The Hamiltonian for an electron in a magnetic field H in the z-direction can be written

$$\mathscr{H} = \frac{1}{2m}\{p_x^2 + (p_y + m\omega_0 x)^2 + p_z^2\} \qquad \ldots\ldots(1)$$

the eigenvalues and eigenfunctions being

$$\epsilon(n, k_y, k_z) = (n + \tfrac{1}{2})\hbar\omega_0 + \frac{\hbar^2}{2m}k_z^2 \qquad \ldots\ldots(2)$$

$$\psi(n, k_y, k_z) = L_y^{-1/2} L_z^{-1/2} \phi_n(x + \lambda^2 k_y) \exp(ik_y y) \exp(ik_z z). \qquad \ldots\ldots(3)$$

In these, $\phi_n(x + \lambda^2 k_y)$ is the wave function for a harmonic oscillator of frequency ω_0 in its nth excited state, k_y and k_z are the wave numbers associated with the y and z directions and L_y and L_z are the corresponding normalization lengths. We have put $\hbar/m\omega_0 = \lambda^2$.

In order to calculate the relaxation time for scattering of the electrons by the acoustic modes of the lattice vibrations, we take the perturbation used by Argyres and Adams (1956), viz.

$$\mathscr{H}_1 = E_1 \sum_{\mathbf{q}} q Q_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}}. \qquad \ldots\ldots(4)$$

We assume that the scattering is elastic and that the temperature is sufficiently high that the number of phonons N_q obeys the relation $N_q \simeq N_q + 1 \simeq kT/\hbar qs$, where s is the sound velocity.

Argyres and Adams assumed that all the electrons were always in the state $n = 0$ and found the relaxation time by calculating the rate of loss of momentum in the z-direction. We shall remove the restriction $n = 0$ and shall verify that, under the above assumptions, a general time of relaxation can be defined. The assumption is inherent in the analysis that $\omega_0\tau \gg 1$, so that the electrons can be considered to be scattered from one stationary state of the Hamiltonian (1) to another.

Using the formalism of Wilson (1953), we find, for the rate of change of the distribution function $f\{\epsilon(n, k_y, k_z)\}$ due to collisions with the lattice

$$\left\{\frac{\partial f(\epsilon)}{\partial t}\right\}_{\text{coll}} = \sum_{n', k_y', k_z',} [f(n', k_y', k_z')\{1-f(n, k_y, k_z)\}W(n', k_y' k_z'; n, k_y, k_z)$$
$$-f(n, k_y, k_z)\{1-f(n', k_y', k_z')\}W(n, k_y, k_z; n', k_y', k_z')] \qquad \ldots\ldots(5)$$

where $W(n', k_y', k_z'; n, k_y, k_z)$ is the probability per unit time of transitions from the state n', k_y', k_z' to the state n, k_y, k_z and is given by

$$W(n', k_y', k_z'; n, k_y, k_z) = \frac{2\pi E_1^2 \boldsymbol{k} T}{\hbar \rho s^2 V} \delta\{\epsilon(n', k_y', k_z') - \epsilon(n, k_y, k_z)\}$$
$$\times \sum_{\mathbf{q}} |<n', k_y', k_z'| e^{i\mathbf{q}\cdot\mathbf{r}} |n, k_y, k_z>|^2. \qquad \ldots\ldots(6)$$

In this, ρ and V are the density and volume of the material. By carrying out the sums over $\mathbf{q}$, k_y', k_z' one can show that

$$\left\{\frac{\partial f(\epsilon)}{\partial t}\right\}_{\text{coll}} = -\frac{E_1^2 \boldsymbol{k} T \boldsymbol{m}^{3/2} \omega_0}{2^{1/2}\pi \hbar^3 \rho s^2} \{f(\epsilon) - f_0(\epsilon)\} \sum_{n'} \{\epsilon(n, k_y, k_z) - (n' + \tfrac{1}{2})\hbar\omega_0\}^{-1/2} \qquad \ldots\ldots(7)$$

where $f_0(\epsilon)$ is the equilibrium distribution function. Thus a time of relaxation exists and is given by

$$\frac{1}{\tau\{\epsilon(n, k_y, k_z)\}} = \frac{E_1^2 \boldsymbol{k} T \boldsymbol{m}^{3/2} \omega_0}{2^{1/2}\pi \hbar^3 \rho s^2} \sum_{n'} \{\epsilon(n, k_y, k_z) - (n' + \tfrac{1}{2})\hbar\omega_0\}^{-1/2}. \qquad \ldots\ldots(8)$$

The sum over n' can be performed using the Poisson sum formula (Dingle 1952) and gives

$$\frac{1}{\tau(\epsilon)} = \frac{2^{1/2} E_1^2 \boldsymbol{k} T \boldsymbol{m}^{3/2} \epsilon^{1/2}}{\pi \hbar^4 \rho s^2} \left[1 + 2^{-1/2}\left(\frac{\hbar\omega_0}{\epsilon}\right)^{1/2} \sum_{p=1}^{\infty} (-1)^p \cos\left(\frac{2\pi p \epsilon}{\hbar\omega_0} - \frac{\pi}{4}\right) p^{-1/2}\right.$$
$$\left. + \left(\frac{\hbar\omega_0}{\epsilon}\right) \sum_{\nu=1,3,5,}^{\infty} \frac{(-1)^{(\nu-1)/2}(2\nu-1)!\, B_{(\nu+1)/2}(2^\nu - 1)(\hbar\omega_0)^\nu}{(8\epsilon)^\nu(\nu-1)!(\nu+1)!}\right] \qquad \ldots\ldots(9)$$

where the B's are Bernoulli numbers.

The time of relaxation therefore shows an oscillatory dependence on the magnetic field analogous to that in the de Haas–van Alphen effect.

To find $\{\partial f(\epsilon)/\partial t\}_{\text{Fields}}$, we use the perturbation $-\epsilon E z$, i.e. we consider only the longitudinal magnetoresistance. Using the method of Jones and Zener (1934), we find the usual expression

$$\left\{\frac{\partial f(\epsilon)}{\partial t}\right\}_{\text{Fields}} = \frac{\boldsymbol{e}E}{\hbar}\frac{\partial f(\epsilon)}{\partial k_z}. \qquad \ldots\ldots(10)$$

For the transverse magnetoresistance, the problem of finding an expression for $\{\partial f(\epsilon)/\partial t\}_{\text{Fields}}$, is more complicated.

The solution of Boltzmann's equation is then, to first order,

$$f(\epsilon) = f_0(\epsilon) + \frac{\tau \boldsymbol{e}E}{\hbar}\frac{\partial f_0(\epsilon)}{\partial k_z} \qquad \ldots\ldots(11)$$

with $$f_0(\epsilon) = \left[\exp\left(\frac{\epsilon - \zeta}{\boldsymbol{k}T}\right) + 1\right]^{-1} \quad \text{(Titeica 1935).}$$

Calculating the conductivity of a metal in the usual way, we then find

$$\sigma = Ne^2\tau(\zeta)\, m^{-1}$$

where ζ is the Fermi level in the presence of the magnetic field, N is the concentration of electrons and the relation between them is

$$\begin{aligned} N &= \frac{2^{1/2} m^{3/2} \omega_0}{\pi^2 \hbar^2} \sum_n \{\zeta - (n+\tfrac{1}{2})\hbar\omega_0\}^{1/2} \\ &= \frac{(2m\zeta)^{3/2}}{3\hbar^3\pi^2}\left[1 + \frac{3}{4\sqrt{2\pi}}\left(\frac{\hbar\omega_0}{\zeta}\right)^{3/2} \sum_{p=1}^{\infty} (-1)^p \sin\left(\frac{2\pi p\zeta}{\hbar\omega_0} - \frac{\pi}{4}\right) p^{-3/2} \right. \\ &\quad \left. + \tfrac{3}{4}\left(\frac{\hbar\omega_0}{\zeta}\right)^2 \sum_{\nu=0,2,4,6,\ldots}^{\infty} \frac{(-1)^{\nu/2}(2\nu-1)!\,B_{(\nu+2)/2}(2^{\nu+1}-1)(\hbar\omega_0)^\nu}{(8\zeta)^\nu(\nu-1)!(\nu+2)!}\right]. \end{aligned}$$

Using a different technique, Zil'berman (1955) has considered the related problem in which the scattering is caused by impurities.

ACKNOWLEDGMENTS

The author would like to thank the Admiralty for permission to publish this note.

REFERENCES

ARGYRES, P. N., and ADAMS, E. N., 1956, *Phys. Rev.*, **104**, 900.
DINGLE, R. B., 1952, *Proc. Roy. Soc.* A, **211**, 500.
JONES, H., and ZENER, C., 1934, *Proc. Roy. Soc.* A, **144**, 101.
TITEICA, V. S., 1935, *Ann. Phys., Lpz.*, **22**, 129.
WILSON, A. H., 1953, *The Theory of Metals*, 2nd edition (Cambridge : University Press).
ZIL'BERMAN, G. E., 1955, *J. Exp. Theor. Phys., U.S.S.R.*, **29**, 762.

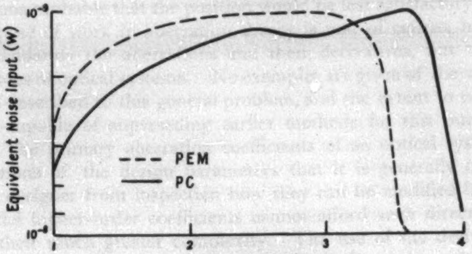
LETTERS TO THE EDITOR

Photoelectric Effects in InAs at Room Temperature

The electrical and optical properties of indium arsenide indicate that it may prove a useful material for the construction of photocells; its maximum sensitivity should lie in the 3–4 μ region, which is particularly interesting for infra-red spectroscopy.

In an initial study we have observed photoconductive (PC) and photoelectromagnetic (PEM) effects in purified InAs at room temperature. The material was n-type with 3×10^{16} donors cm^{-3}, and an electron mobility of 24 000 cm^2 v^{-1} sec^{-1}. The carrier lifetime, deduced from the ratio of the photoconductive to photoelectromagnetic voltages, was 0·2 μsec. The ingots used were polycrystalline, but since the crystallites were fairly large it was possible to cut specimens without grain boundaries in them. When grain boundaries were included, pronounced photovoltaic effects could be observed. Photovoltaic effects have been reported previously in InAs cooled to 77°K (Talley and Enright 1954).

Photoconductive and photoelectromagnetic cells were constructed by etching specimens 2×1 mm to a thickness of about 50 microns. The resulting sensitivity and spectral response are shown in the figure. (The equivalent noise input is defined as the radiation intensity necessary to give a voltage equal to the noise voltage in a 1 cycle bandwidth.) The sensitivity level is comparable with that of available PbSe and InSb cells (Starkiewicz 1948, Hilsum and Ross 1957). The energy gap deduced from the photoconductive cut-off is 0·33 ev, in good agreement with determinations from optical and electrical data (Oswald and Schade 1954, Weiss 1956).



Spectral sensitivity of InAs photocells. Area = 2 mm^2.

The use of purer material with a longer lifetime is expected to yield cells of higher sensitivity, and the preferred mode of operation will then probably be photoconductive or photovoltaic. This is in contrast with indium antimonide, where the photoelectromagnetic mode gives the simplest practical photocell.

The material used in these experiments was prepared by Mr. J. T. Edmond and Mr. W. Harding. Details of the techniques used in the preparation will be given in a future publication.

Acknowledgment is made to the Admiralty for permission to publish this Letter.

Services Electronics Research Laboratory,
Baldock, Herts.
24th July, 1957.

C. Hilsum.

Hilsum, C., and Ross, I. M., 1957, *Nature, Lond.*, **179**, 146.
Oswald, F., and Schade, R., 1954, *Z. Naturforsch.*, **9a**, 611.
Starkiewicz, J., 1948, *J. Opt. Soc. Amer.*, **38**, 481.
Talley, R. M., and Enright, D. P., 1954, *Phys. Rev.*, **95**, 1092.
Weiss, H., 1956, *Z. Naturforsch.*, **11a**, 131.

REVIEWS OF BOOKS

Optical Aberration Coefficients, by H. A. BUCHDAHL. Pp. xx + 336. (London: Cumberlege; Oxford: University Press.) 1954. 50*s*.

This book describes a quite new approach to the aberration theory of optical systems. Using elegant mathematical methods, iterative formulae are derived which make possible the calculation not only of the primary aberration coefficients which are commonly used in lens designing, but also the second-order ones, and in principle those of any order, and in addition the first and higher derivatives of these coefficients with respect to changes in the design parameters. The mathematics is not inherently difficult, though the presentation is that of a mathematician rather than a physicist, and it is extremely voluminous; appendix J, which lists the more commonly used symbols, occupies eight pages. A number of numerical calculations is given on a Cooke triplet lens of relative aperture $f/5{\cdot}6$ designed to cover a field of ± 18°.

The amount of computing work in evaluating the aberration coefficients is considerable, increasing rapidly as successively higher orders are considered, so that the number of orders required to give a reasonable approximation to the total aberration is of basic importance; this rate of convergence depends of course on the particular optical system. For the triplet lens the total aberration has been calculated using normal ray-tracing methods, and where three orders of aberration have been calculated these show good agreement with the total aberration; for other aberrations two orders only are given, and here in the worst cases the difference between the total aberration and that given by the first two terms is less than the second-order term. For lenses of considerably larger aperture or field, which constitute the main problems of present-day optical design, it seems probable that the position would be less satisfactory.

The purpose of work in aberration theory is not, of course, merely to find means of calculating the aberrations and their derivatives, but to reduce the total aberrations of optical systems. No examples are given of the application of the methods described to this general problem, and the extent to which they are likely to be capable of superseding earlier methods for this purpose appears debatable. The primary aberration coefficients of an optical system are such simple functions of the design parameters that it is generally obvious to an experienced designer from inspection how they can be modified in any desired direction; the higher-order coefficients cannot afford such direct guidance on account of their much greater complexity. The use of the derivatives of the aberration functions appears to be inherently limited; as the author points out, any complete solution using all the derivatives of several orders is quite impossible, so that generally first derivatives of the lower order coefficients only would be used, possibly with a few selected higher ones. Since methods are already available of determining the first derivatives of the total aberration (determined by ray-tracing), the advantages of the new methods appear somewhat uncertain. In one of the notes added in proof the author states that, since writing the book, his methods have been used in his laboratory for designing a variety of optical systems. If some details of this work had been included in the book, the value of the methods would be easier to assess.

The book is generally well produced. It would be easier to read if more diagrams had been provided, and page references given in the text to the few diagrams that do occur.

C. G. W.

Handbuch der Physik, Vol. 21, *Gas Discharges* I; *Electron emission*, edited by S. FLÜGGE. Pp. viii + 683. (Berlin: Springer, 1956.) DM. 132.

This is the first of two volumes in the *Handbuch* dealing with gas discharges and associated topics. It is concerned entirely with the fundamental processes in the gas or on the surfaces. The general standard of the articles is excellent, and the authors have accumulated a vast amount of widely scattered information and presented it in a coherent and readable form which makes this a most valuable book. In a few places it falls short of the ideal of a *Handbuch* (i.e. a source in which one can find everything somewhere).

The first article by W. B. Nottingham on Thermionic Emission deals exclusively with emission of electrons. After a brief historical survey, and classification of types of emitting surface, the theory is developed rigorously, especially that relating to complex and oxide coated cathodes, where new work is incorporated. Modern emitters (e.g. 'dispenser' cathodes) are mentioned and there are many numerical tables of useful data. Four pages filled with a glossary of symbols might have been better used to refer to more experimental work. An article by R. H. Good and E. W. Muller on Field Emission begins with an account of the main discoveries and the simple physical picture of the process. The Fowler–Nordheim theory is extended to describe emission at increasing temperatures until it becomes thermionic. Experimental results are summarized, showing how the current and velocity distribution of electrons depend on field, temperature and work function. Emission from semiconductors, and the use of modern pulse techniques are mentioned. The field emission microscope and the ion microscope are described and their uses illustrated with many beautiful photographs.

There is only one article written in German, a comprehensive treatment (with 400 references) of Secondary Electron Emission by R. Kollath. He describes the methods and the results of many experiments in tables and graphs. These show how the total yield and distribution of the secondary electrons depend on the energy and angle of the primaries, the material, temperature and crystal structure of the surfaces. G. L. Weissler writes lucidly on photoionization in gases and photoelectric emission from solids. He gives the basic theory and experimental methods, and presents the results of all the measurements in consistent form, the absorption (expressed as μ and σ) being plotted against the wavelength of the light (expressed in Å and ev). There are valuable curves of photoelectric yields and electron energy distributions, and a discussion of contaminated surfaces.

There follows an account by W. P. Allis of the Motion of Electrons and Ions in Gases. Rigorous and most elegant mathematical treatments describe their drift velocities, energy distributions etc., when subjected to electric and magnetic fields. It is disappointing then to find no numerical conclusions, no mention of any experimental work or measured values to compare with theory. This omission leaves an obvious gap which is only partially filled by a table of drift velocities in another article.

Two short articles by L. B. Loeb are devoted to Negative Ions and Recombination. The various processes by which an electron can attach itself to a molecule are set out comprehensively in table form, and the measurements of the probability of this attachment and the energy required critically discussed.

A similar table shows the many possible types of recombination, and gives up to date values of the coefficient.

A. von Engel writes clearly about Ionization by Electrons in electric and magnetic fields of varying geometry. He discusses the measurements of α, its connection with the ionization efficiency and the energy distribution of the electrons, and the effect of gaseous impurities. It is interesting to see that an electronegative addition often causes α to increase. There are many valuable graphs and tables.

The book ends with an excellent account of Secondary Effects by P. F. Little, who after defining a 'secondary process' shows how a given process can be primary or secondary depending on circumstances. Ionization caused by ions, by photons, and by excited, metastable and fast neutral atoms and other factors, is discussed and experiments described which enable one to identify their contributions to γ. The presentation is clear and systematic.

A common feature of most articles is the emphasis on the need for the greatest cleanliness of gases or surfaces. Modern hygienic measurements have often falsified older values; I suppose it is heresy to wonder whether further refinements will restore the *status quo ante*.

The editor and publishers are to be congratulated not only for their superb production of the book, but also for flouting the tyranny of Standardizing Committees. The printer's function here is evidently simply to convey the meaning of the author, who is allowed to choose his own symbols, with the result that even familiar equations appear in recognizable form. How refreshing.

G. FRANCIS.

.B
0
7

Micro-Indentation Hardness Testing, by B. W. MOTT. Pp. viii + 271. (London: Butterworths Scientific Publications, 1956.) 42*s*.

If indentation tests at loads of the order of 1–100 g could be made with the same facility and reproducibility as conventional hardness measurements, those concerned with the structure and mechanical properties of solids would possess a technique of great power. In addition to obvious metallurgical uses, such a test could be applied to many non-metallic materials: plastics which creep extensively, or minerals which are brittle, around ordinary hardness indentations.

Unfortunately, this ideal is far from being realized, for there are at least three formidable difficulties. As the load is lowered, the test becomes increasingly susceptible to effects associated with the inertia of the moving part of the apparatus, particularly from small vibrations. Errors of measurement arising from the constant resolution of the microscope, and the difficulty of locating exactly the corner of the impression, become proportionately greater as the size of the impression is reduced. Finally, it is very uncertain how far the inherent properties of very small volumes of material, particularly at the surface, represent bulk values.

Mr. Mott's painstaking survey of the achievements and hazards of microhardness work—or 'micro-indentation' as he prefers, not without reason, to call it—is therefore a book well worth reading by anybody likely to use one of the eighteen types of instruments described in his second chapter. The contributions of various factors to the increase in apparant hardness with decreasing load, and to the variation of logarithmic index, are assessed, and the relations of the various hardness scales to each other and to the mechanical properties of the material, are clearly developed. The scope of the technique is indicated by the many kinds of investigation described, and there are some valuable tables of the hardness (Vickers and Knoop) and logarithmic index, at various loads, of numerous metallic and non-metallic materials. In short, sufficient evidence is given to enable the reader to decide for himself how far low load hardness tests are to be relied on, and what improvements in current practice might be made.

Two questions seemed to me to deserve much more attention than they have yet received. First, there exists no adequate theoretical analysis of the plastic behaviour of small volumes of material to compare with experiment. Mr. Mott does make an attempt at this, based on a model of Orowan's, but I think the problem is much more complicated than he indicates. Some effects which are at present ascribed to instrumental errors may be, in fact, a property of the specimen; for instance, the effect observed by Whitaker (p. 70) that the hardness increases to a maximum as the depth of indentation is reduced, then decreases rapidly. This decrease is put down to loading errors; but one might expect such a drop in hardness when plastic flow is confined to the surface layer, where dislocation theory predicts a yield stress about half that of the interior.

Secondly, I suggest that many of the results, particularly in the lower range of loads, are vitiated by the lack of statistical evaluation. The few examples given encourage the suspicion that single hardness figures are very difficult to interpret below about 25 g load—the table on p. 97 shows that the standard deviation is then already of the order of 10% of the mean, and goes up rapidly as the load is decreased. The quantitative value of a test would be greatly improved if results were always given in terms of a mean hardness accompanied by some sort of confidence limit; the dispersion may be in part a significant property of the material. Until more figures of this sort are available, and some of the important sources of variability (such as surface preparation) are standardized, tests at very low loads are best regarded as convenient qualitative methods, like the Mohs' scale, of ranking the mechanical properties of solids.

K. E. PUTTICK.

Spheroidal Wave Functions (*including tables of separation constants and coefficients*), by J. A. STRATTON, P. M. MORSE, L. J. CHU, J. D. C. LITTLE and F. J. CORBATÓ. Pp. xiii + 613. (New York: John Wiley, and Technology Press of M.I.T., jointly; London: Chapman and Hall, 1956.) £5.

We are still far from having adequate tables of all special functions which arise when the scalar wave equation is separated in various three-dimensional coordinate systems. Wave functions associated with the rectangular box, the sphere and circular cylinder are fairly well tabulated; those for the elliptic and parabolic cylinders are either well advanced or being actively pursued, and the spheroids, prolate and oblate, are certainly comparable with these in importance.

This volume represents a stage—but only a stage—in the provision of spheroidal wave functions, and to those who have already benefited from the earlier tables of Stratton, Morse, Chu and Hutner, it is a welcome reminder that this effort, initiated some 25 years ago, is continuing. The task is a major one, because in addition to the continuous variables there are three parameters to consider, two integral parameters l and m (which in the language of quantum mechanics are the angular momentum and magnetic quantum numbers) and the continuous parameter h (or g) representing the number of wavelengths in the interfocal distance of the prolate spheroid, or in the focal diameter of the oblate spheroid.

The stage reached is that in which the regular radial wave function is expanded as a series in spherical Bessel functions, and the angular function in terms of Legendre functions. The coefficients in both series, and values of the mutual separation constant, are tabulated to 7 significant figures, for $l, m = 0(1)8$ $(l \geqslant m)$ and h (or g) $= 0(0{\cdot}1)1(0{\cdot}2)8$. Thus these coefficients, combined with existing tables of spherical Bessel functions and Legendre polynomials, offer scope for numerous applications in physical problems. Nevertheless, it is expected (and to be hoped) that actual tables of spheroidal wave functions will become available in the future.

A feature of the present production is that both calculation and printing were carried out as a single process on the Whirlwind computer at the Massachusetts Institute of Technology. This gave rise to obvious technical problems in planning the printing layout of the tables, but the result is reasonably satisfactory. However, a more detailed index to the tables would have been useful.

It is an interesting comment on modern developments that most of the programming was completed in six months, and the time of production on the machine was about ten hours. R. A. BUCKINGHAM.

Rheology: *Theory and Applications*, Vol. 1, edited by F. R. EIRICH. Pp. xiv + 761. (New York: Academic Press; London: Academic Books, 1956.) $20.00; £7 3*s*.

This is the first of three volumes of a work which is to cover most of the field of experimental and theoretical rheology. It is claimed that it may well be the forerunner of a unified theory. While this seems unlikely, the first volume, nevertheless, brings within a single cover many authoritative essays on fundamental rheological theory.

Most rheological tests are at present of the 'vector' type, i.e. a test-piece is subjected to stresses or strains in a single direction, any stresses (and even strains) in other planes being generally ignored. Such procedures are proving increasingly inadequate and a tensorial basis for rheological theory is essential. M. Reiner's chapter on the tensor theory of macro-rheology provides a stimulating introduction to this approach.

Plastic phenomena in metals and other crystalline systems, and crystal imperfections, are competently dealt with by a number of leading authorities. High-pressure work is treated by R. B. Dow, theories of viscosity by A. Bondi, and large elastic deformations by R. S. Rivlin. There is a chapter on visco-elastic behaviour and several good essays on macromolecules in solution. Birefringence is discussed by A. Peterlin, non-Newtonian flow by J. G. Oldroyd, and acoustic methods by R. B. Lindsay.

While there is little in this volume which is not already in other textbooks, it is claimed, with some justice, that it is useful to have this range of essays within a single work. On the other hand, most rheologists will have access to the earlier writings of these distinguished authors; and, while this new treatise is rather expensive for most individuals to buy, one wonders whether it will be of any great advantage for libraries to acquire so much of the old material in this new form.

Nevertheless, the book is, in its way, a monumental work. It has also the merit of being excellently produced and printed. Your reviewer could find only one major misprint: the name Treloar is misspelt in several places.

It is to be hoped that the second and third volumes, while maintaining the very high standard of the first, will include rather more material not already available in standard textbooks. G. W. SCOTT BLAIR.

Viscous Flow Theory—I. Laminar Flow, by SHIH-I PAI. Pp. xvi + 384. (Princeton: Van Nostrand; London: Macmillan.) 58*s*.

This book presents an account of those parts of the theory of viscous motion that are likely to be of greatest interest to engineers, particularly aerodynamicists. Even within this defined field, some degree of selection has necessarily been made by the author to keep the size of the book within reasonable limits, and a middle course is steered between the demands of the undergraduate student requiring a broad survey of fundamentals and those of the research specialist who seeks a more thorough and detailed development of his particular topic.

The book opens with a brief description of a few thermodynamical concepts and of the classical kinetic theory of gases which introduces the notion of viscosity. The relations between stress and rate of strain are set up in the usual manner from which is derived the momentum equation. This discussion is expressed in terms of rectangular cartesian components; little use is made of the simpler and more concise vector notation. Some exact solutions of the Navier–Stokes equations are obtained including Squire's jet and Morduchow and Libby's description of the structure of a shock wave. There is a chapter on the mathematical properties of the Navier–Stokes equations but the need for brevity has resulted in its being rather unsatisfying for mathematically inclined readers and probably very difficult for those who are not. After a chapter on low Reynolds number approximations, describing the classical Stokes–Oseen investigations and, briefly, lubrication theory, the remainder of the book is devoted to boundary layer theory. This is probably the best section of the book and the discussion is as detailed as any outside such works as Schlichting that are devoted entirely to the subject. Exact and approximate methods of solution of the two-dimensional boundary layer equations are described quite fully, together with the theory of axially symmetrical boundary layers and flows with suction. Questions of stability are touched rather lightly and the book concludes with a chapter on the so-called linearized theory of viscous compressible motion. We are promised a second volume on turbulent flow.

The principal impression of this work is that it places considerable emphasis upon formal manipulation of equations rather at the expense of a genuine understanding of the physical mechanisms involved. The reviewer was occasionally conscious of lack of precision in the argument and of statements that a student might find misleading. On the whole, however, the book is likely to be worth while to those who do not wish to penetrate too deeply into any particular aspect of the subject. O. M. PHILLIPS.

Laboratory Glassblowing, by L. M. PARR and C. A. HENDLEY. Pp. 160. (London: George Newnes, 1956.) 21*s.*

The authors of this small volume of 160 pages set out to give instruction in the fundamental operations of glassblowing and glass working techniques. The hope is expressed that laboratory workers of reasonable dexterity will be enabled to acquire sufficient skill to make repairs and fashion quite complicated glass apparatus. The basic processes in 'scientific glassblowing' are few in number and, individually, are easy to execute after adequate practice. Craftsmanship begins to show itself when the student can plan the construction of a complex piece of work as a sequence of elementary steps which may be carried out in proper order without at any time building up harmful strain in the glass. Skill is reflected in the quality of the work and comes, sometimes suddenly, with mastery of the art of manipulating glass in as fluid a state as possible while still retaining control over its movements. These aspects of the practical side of glassworking are repeatedly stressed by the authors and they provide the necessary instruction in well-established procedures by way of detailed description and stage-by-stage illustrations. The student is shown how to draw working 'points', blow bulbs and flange tubing; and how to make butt-joints, T-joints, bends, double-seals and spirals. Several ways of cutting tubing are described and the importance of working with clean material is not overlooked. Surprisingly, the useful technique of flame-cutting is not included.

Almost the whole of two chapters is allotted to the fabrication and grinding of stopcocks because, it seems, "occasions will always arise when the one you want is not available, and it has to be made". It is a most interesting subject to read about but, on economical grounds, the making of stopcocks should surely be left to the specialist. There are stockists who carry a wide range of standard products and the recognized manufacturers are most willing to make up special patterns. The laboratory worker would have been better served, perhaps, if the space had been devoted to the shaping and forming of non-standard parts by simple grinding operations as in the preparation and fitting of stoppers, ball-valves and flanged joints and to the use of the tube-cutter for making holes and circular blanks. One needs guidance but very little skill or training to fulfil *ad hoc* requirements of this kind which are properly within the province of the laboratory technician.

The introduction to a craft would be incomplete without a background of accurate, well proved technological information. In this book there are chapters dealing with equipment, blow-pipes, glass and glass-annealing, glass-to-metal seals and special techniques such as the graduating of vessels; but in several respects the treatment is disappointing. Thus on p. 12 we are told how to feed a mixture of oxygen and compressed air to a blow-pipe. If the air is supplied from a mechanical compressor and the several stop-cocks are not turned off in the correct order, an explosion could occur in the compressor in spite of non-return valves being fitted. The oxygen should either feed directly into the flame or be entrained in the air-stream from a source at atmospheric pressure. On p. 22 it is stated that when glass will take on a sooty deposit during flame annealing, it is cool enough to set aside. This is not true for lead glass where the upper annealing temperature is below the sooting temperature. Glass technologists will be interested in the remarks on devitrification (p. 30).

Chapter 8 on metal-to-glass seals should be approached warily. In spite of the wealth of published data on practical glass–metal combinations those of the authors' choosing are sometimes not fully specified or are technically unsound. For example it *is* possible to make seals between Pyrex glass and tungsten, molybdenum or platinum with special cooling schedules (as distinct from proper annealing) but the technician should not be given unnecessary and unrewarding exercises of this nature when other, more suitable glasses are available for each of these metals. This confusion between preferential cooling and annealing is again apparent in the discussion on nichrome–soda glass seals. Eldred's purpose in devising Dumet wire is also not fully appreciated. One would like to know which steel has an expansion coefficient of $10{\cdot}5 \times 10^{-6}$ (p. 95) and which borosilicate glass (among many) can be joined to porcelain or mullite (p. 112).

Mismatched copper–glass seals are discussed at some length along conventional lines. It is surprising how often we find them described in scientific publications as 'Housekeeper' seals although the name of the pioneer was W. G. Houskeeper. This book commits the same error.

Attention is drawn to the following slips in proof-reading. The ordinates of figure 3.1 are incorrect. On p. 25 under Wembley Neutral glass "linear coefficient of linear expansion" should read ". . . . of thermal expansion". On p. 26 under Hysil glass "linear annealing temperature" should read "lower annealing temperature". The two glasses H.H. and W.1 described on p. 28 should be placed at the end of the section on borosilicate glasses (p. 27) and before the aluminosilicate glass.

C. H. SIMMS.

Advanced Calculus, by R. C. BUCK. Pp. viii + 423. (New York: McGraw-Hill, 1956.) (International Series in Pure and Applied Mathematics). 64*s*.

In this valuable book the author treats with care a number of subjects which are not easily accessible elsewhere, as well as a certain amount of more routine material. In the systematic account of Riemann integration, transformations of Euclidean space and their inverses and change of variable in integration and differentiation, the emphasis is always on several variables. Line and surface integrals are dealt with and there is an introduction to differential geometry, vector analysis and the theorems of Green, Gauss and Stokes, which are treated by means of the formalism of Cartan's exterior differential calculus. The book finishes with a short account of the calculus of variations and a 6-page appendix on the real number system.

Further topics treated include elements of limits, mean value theorems and l'Hôpital's rule, set theory and the study of continuous functions on compact and connected sets, and theory of series including power series and uniform convergence, but not Fourier series.

The treatment is rigorous and reasonably complete. In the more difficult cases the author refers to other monographs from time to time or gives the proof only in a simple special case. This technique is probably inevitable when dealing with, for instance, the complexities of Stokes' theorem. On the other hand it may be worth while in a future edition to include a proof that a linear transformation $T(x)$ in Euclidean space multiplies volumes by det T instead of referring the reader to an algebra textbook. For this result is used as a key theorem in all the subsequent work on transformations of multiple integrals in

Chapter VI. Consequently, the reader is left in some doubt as to what has really been proved.

However, this is a minor criticism and is in any case a matter of taste. This book will be valuable for mathematicians and scientists who want the formulae and techniques of advanced calculus and are prepared to face some mathematical difficulties in order to see things done properly. There are many excellent examples to illustrate the text. These are starred in accordance with their difficulty. The ordinary one would be suitable for undergraduates in mathematics, and two star examples for exceptionally brilliant men. However, hints of solutions are given. There is an index and a list of suggested books for further reading at the back. The printing is excellent. W. K. HAYMAN.

Lecture Notes on the Use of the Microscope, by R. Barer. 2nd Edn. Pp. vii + 76. (Oxford: Blackwell Scientific Publications, 1956.) 7*s*. 6*d*.

This very useful little book has now appeared in a second edition, and is still devoted to a clear exposition of the visual use of the ordinary microscope. It does not deal with photomicrography, nor with such topics as phase-contrast ; it is intended for the beginner, and he will find it a trustworthy guide.

It is probably true that the standard of production of microscope objectives is now good enough—or could be good enough—to obviate much of the tube-length adjustment, etc., in which the microscopists of an earlier generation took so much interest. The modern worker may buy his outfit complete ; if he goes to a reputable firm he may expect satisfactory ' optics ', and need take no more precautions than those which are so well described in the present handbook. If, however, he must use apparatus of doubtful vintage he will do well to acquire some knowledge of the testing methods which are described in the fuller textbooks, though indeed they are hard to learn adequately from any book and are justifiably omitted from these excellent ' lecture notes' ! L. C. M.

Progress in Metal Physics, 6. Edited by B. Chalmers and R. King. Pp. viii + 354. (London and New York: Pergamon Press, 1956.) 70*s*.

The sixth issue well maintains the high standard established by this series, which comprises reviews written in general by authors who are not only learned in their subjects but have made substantial individual contributions to them. Two of the papers (Mehl and Hagel on the Austenite : Pearlite Reaction, and Sully on Creep) deal with classical metallurgical subjects. Lawson's article on the effect of hydrostatic pressure on electrical resistivity in metals brings out a complexity of phenomena which one might not have suspected. Hardy dealing with filamentary growth of metals puts a fashionable subject into historical perspective. The present reviewer has read with particular interest Knacke and Stranski on the Mechanism of Evaporation, and Hirsch on the Mosaic Structure. The former article brings together a large amount of information, more about non-metals than metals : but to be thoroughly up to date on the subject the reader should also pay attention to a subsequent paper on the evaporation coefficient by Littlewood and Rideal (*Trans. Faraday Soc.*, 1956, **52,** 1598) which reassesses the cooling correction and shows that many evaporation coefficients are much nearer to 1 than has been supposed.

Hirsch's paper (104 pp. with 393 references) is surely the completest account of its subject yet presented. The subject gradually acquires precision. It is even possible that it is no longer pedantic to use the conversion factor from number of dislocation lines intersecting a plane to line length per unit volume: but this factor is not 3 as stated on p. 252 and employed throughout the paper, but 2—i.e. the reciprocal mean cosine of the angle between spatially random directions and a chosen axis. The corresponding factor for lines random in a plane is $\pi/2$, not 2. F. C. F.

The International Dictionary of Physics and Electronics, Senior Editor W. C. MICHELS. Pp. xii + 1004. (London: Macmillan, 1957.) 120*s*.

This is a wholly admirable work, for which the highest credit must go to the Senior Editor, Professor Walter C. Michels, of Bryn Mawr College, and his Board of fourteen, of whom four, Mendelssohn, Townsend, Yoffe and Ziman are British. Until fairly recently, physics (as contrasted for instance with medicine or with mineralogy) was very economical in its terms, but today there will be few people who can read through the index of a number of the *Physical Review* without stumbling over a few words such as ' plasmons ', ' plasmoids ' or ' spallation '. The newcomer, electronics, was generous right from the start in its terminology, which has blossomed out in terms such as phantastron, phasitron, donutron and recently the wamitron.

Now at last we have a tome which on its 1004 pages contains almost everything, from Isotopic Spin to Insertion Loss. Confronted with a book which contains on almost every page at least one word which is new to him, the natural reaction of the reader will be to ' try to beat the book ' to save his self-esteem. I must confess though that apart from some (not quite fair) combinations such as Bitter Domains and Stieltjes Integrals, I could beat the book only with two terms which I have coined myself, the Logon and the Hologram. I found in it the third of my terms, the Kipp Relay (which is widely attributed to a fictitious Herr Kipp).

This book ought to be in every laboratory library, and in the private library of every physicist or electronic engineer who can afford it. If it has one disadvantage, it is that it weighs more than 5 pounds, and is not a pocket edition which one could consult under the table. D. GABOR.

Digital Calculating Machines, by G. A. MONTGOMERIE. Pp. 262. (London and Glasgow: Blackie, 1956.) 30*s*.

This book is intended as an introduction to calculating machines of the digital type, which the author treats as follows: Chapter I (Introduction), Chapters 2–7 (Desk Calculating Machines), Chapter 8 (Punched Card Machines), and Chapters 10–12 (General Purpose Automatic Machines).

The first chapters give a very complete account of the principles of operation of all types of desk calculating machines including adding and listing machines, lever-set hand-operated machines such as the Brunsviga, key-set hand machines such as the Facit, and electrically operated machines such as the Marchant. The use of these machines for such operations as differencing, summation of products, synthetic division, and other calculations of a scientific and statistical nature are described. For this purpose a symbolic notation is employed to describe the operations of setting a number on a register, depression of the 'motor keys', etc. This also forms a convenient introduction to the subject of 'programming' for large-scale automatic machines, which is dealt with in the last chapter.

The chapter on punched card machines describes the principles of operation of both Powers–Samas and Hollerith equipment and the use of such equipment for solving linear simultaneous equations.

Chapter 9 describes how electromechanical relays and other items of telephone switchgear have been adapted to serve as computing elements in calculating machines, for example, the special purpose machine devised by MacEwan and Beevers (1942) for the summation of Fourier Series. Such machines are now, however, mainly of historical interest.

The chapters on large-scale machines deal mainly with the earlier machines such as the Harvard Mk. I and the ENIAC which serve to explain the principles of storage and number representation. In the last chapter the principles of 'programming' (that is, the preparation of instruction schedules for machines) are explained with the help of an instruction table for a hypothetical machine. This serves to introduce the reader to such ideas as the counting of cycles of operation and the use of sub-routines. R. A. BROOKER.

MacEwan, D., and C. A. Beevers, 1942, *J. Sci. Instrum.*, **19**, 150.

CONTENTS OF SECTION A

PAGE

Prof. R. Eisenschitz and Mr. P. Dean. Electronic Levels of a Model of Liquid Potassium . . . 713

Dr. R. J. N. Phillips. A Phase-Shift Analysis of Nucleon–Nucleon Scattering at 95 Mev . . . 721

Prof. J. M. Cassels, Miss M. Rigby, Mr. A. M. Wetherell and Dr. J. R. Wormald. Radiative Beta Decay of the Pion . . . 729

Mr. A. Ashmore, Mr. R. G. Jarvis, Mr. D. S. Mather and Dr. S. K. Sen. The Energy Spectrum of the 350 Mev Neutron Beam from the Liverpool Synchrocyclotron . . . 735

Mr. A. Ashmore, Mr. R. G. Jarvis, Mr. D. S. Mather and Dr. S. K. Sen. Nuclear Total Cross Sections for 350 Mev Neutrons . . . 745

Research Notes:

Dr. A. L. Stewart. Wave Functions for Helium and Similar Atomic Systems . . . 756

Dr. D. M. S. Bagguley, Dr. J. A. Powell and Mr. D. J. Taylor. Cyclotron Resonance over a wide Temperature Range . . . 759

Dr. E. H. Bellamy, Mr. R. W. P. McWhirter and Mr. P. Palit. The Elastic Photoproduction of Neutral Pions from Helium . . . 763

Mr. P. J. Wood and Prof. G. S. Rushbrooke. On the High Temperature Susceptibility for the Heisenberg Model of a Ferromagnetic. . . . 765

Letters to the Editor:

Dr. R. D. Connor and Mr. I. L. Fairweather. The β-Spectrum of Sulphur 35 . . . 769

Corrigendum:

Prof. J. A. V. Fairbrother, Mr. D. G. Parkyn and Mr. B. M. O'Connor. The K Auger Yield for Tin . . . 770

Reviews of Books . . . 771

Contents of Section B . . . 776

PROC. PHYS. SOC. VOL. 70, PT. 10—B (A. GOSWAMI AND Y. N. TREHAN)

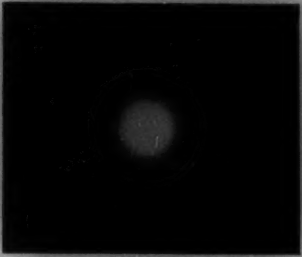


Figure 1. Polycrystalline cupric oxide.

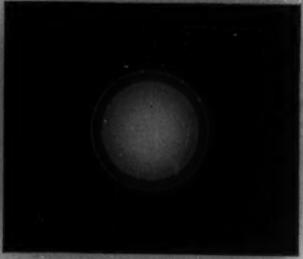


Figure 2. Polycrystalline cuprous oxide.

Figure 3. One-degree {110} orientated cupric oxide.

Figure 4. One-degree {021}+{010} orientated cupric oxide.

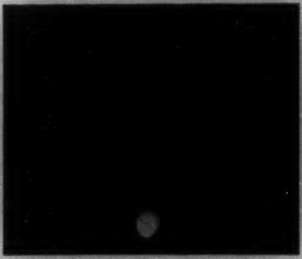


Figure 5. One-degree {110} orientated cuprous oxide.

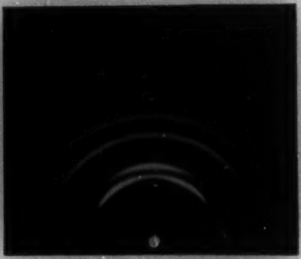


Figure 6. One-degree {100} orientated cuprous oxide.

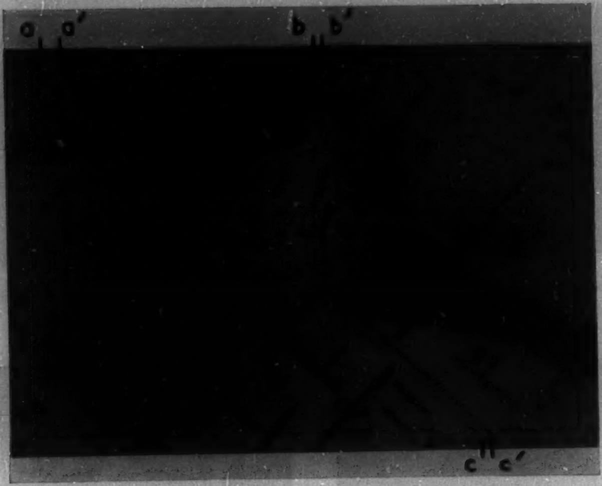


Figure 1. Wall doubling.

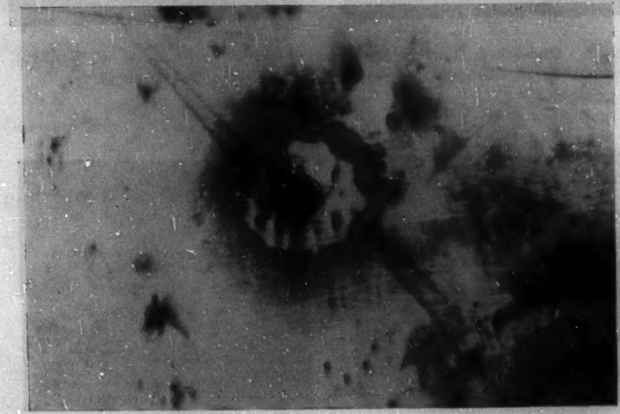


Figure 2. Stain at surface inclusion.

The Reaction of Oxygen with Lead Selenide

By R. H. JONES
Royal Radar Establishment, Malvern, Worcs.

MS. received 18*th April* 1957, *and in revised form* 22*nd July* 1957

Abstract. Measurements are described of the oxidation of lead selenide, in the form of a powdered single crystal, at various temperatures in the range from room temperature to 280°c. These were aimed at gaining fundamental information about the photoconductive sensitization of evaporated layers of lead selenide by oxygen. Results obtained indicate that two processes occur during oxidation. In the range of temperature considered the predominating process appears to be the diffusion of oxygen ions to vacant selenium sites in the lattice; this was found to have an activation energy of 17 kcal mol^{-1}. The second process reveals itself only at the highest temperature employed, and is believed to determine the oxidation rate when all or nearly all the selenium vacancies are occupied by oxygen ions. The postulation of two oxidation processes is supported by earlier electrical measurements on lead selenide evaporated layers exposed to oxygen. Theories of photoconductivity in evaporated lead selenide layers sensitized by oxygen are discussed in relation to these experiments.

§ 1. INTRODUCTION

A PREVIOUS publication (Jones 1957) has described the changes in electrical resistance produced by exposing evaporated lead selenide layers to oxygen. The present work is concerned with an examination of the kinetics of the oxygen–lead selenide reaction. Measurements were made on a single sample of lead selenide of mass 3·8 g prepared by finely grinding a p-type single crystal of material grown in this laboratory. After initially reducing the sample by heating in purified carbon monoxide and outgassing, the sample was exposed to oxygen at a pressure of almost exactly 150 mm in each case at temperatures in the range from room temperature to 280°c, and the quantity of oxygen taken up measured at appropriate time intervals. Following each oxidation the sample was reduced by carbon monoxide and outgassed, the carbon dioxide obtained being collected and measured. Also measurements were made of the surface area of the sample available for reaction, to allow for changes in surface area which may have taken place. The method employed was the physical adsorption of argon at liquid oxygen temperature.

§ 2. DESCRIPTION OF APPARATUS

The gas-handling system used for surface area measurements and oxidation experiments is shown schematically in figure 1. This was constructed from 1 mm bore boro-silicate glass capillary tubing, and employed stopcocks having the same bore. Measurements of space volumes in the system were made using the previously calibrated bulbs of a gas burette as standards.

Oxygen and argon used were of the 'Spectroscopically Pure' grade supplied by the British Oxygen Co. Carbon monoxide was prepared in the laboratory by the action of sulphuric acid on warmed formic acid, both liquids having previously been outgassed. Purification was effected by passing the gas through potassium hydroxide, a liquid-oxygen trap, oxygen-free copper foil at 500°c, anhydrous magnesium perchlorate, potassium hydroxide and a second liquid-oxygen trap.

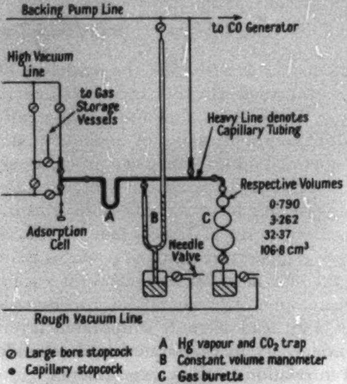

Backing Pump Line
to CO Generator
High Vacuum Line
to Gas Storage Vessels
Heavy Line denotes Capillary Tubing
Respective Volumes
0·790
3·262
32·37
106·8 cm³
A
B
C
Adsorption Cell
Needle Valve
Rough Vacuum Line
⊘ Large bore stopcock
● Capillary stopcock
A Hg vapour and CO₂ trap
B Constant volume manometer
C Gas burette

Figure 1. Gas handling system.

§ 3. SURFACE AREA MEASUREMENTS

Eighteen separate adsorption isotherms were obtained during the investigation. In each case the volume of argon at S.T.P. required to form an adsorbed monolayer on the sample was obtained by estimating the position of the 'B' point on the isotherm (Brunauer 1944). Also the experimental data was plotted according to the relationship for multi-layer adsorption proposed by Brunauer, Emmett and Teller (1938)

$$\frac{p}{p_0 v(1-p/p_0)} = \frac{1}{v_m C} + \frac{p(C-1)}{v_m C p_0},$$

where v is the volume of gas adsorbed at pressure p, p_0 the saturated vapour pressure of the gas at the temperature of the adsorbent, v_m the volume required to form a monolayer and C is a constant related to the heat of adsorption. Over the range of p/p_0 employed, namely about 0·02 to 0·3, an approximately linear plot was obtained in all cases. Both methods of analysis yielded values of v_m which were approximately constant throughout, indicating that the surface area of the sample did not change appreciably. Using the value $14{\cdot}4 \times 10^{-16}$ cm² for the molecular cross-sectional area in the adsorbed argon film, this being the value appropriate to the liquid phase of argon, the value $6{\cdot}5 \times 10^{3}$ cm² was obtained for the selenide surface area. Now the lead selenide crystal is face-centred cubic in symmetry and is consequently most likely to cleave across its cubic faces and produce rectangular-sided particles by grinding. Assuming the powder to consist of equal-sized cubic particles of side d and using the value 8·2 g cm⁻³ for the density of the bulk crystal, the appropriate value for d was found to be about 4 μ. This estimate was later confirmed by microscopic examination of the powder.

§ 4. The Initial Reduction of the Sample

The initial reduction was carried out in stages, the quantity of carbon monoxide taken up and the quantity of carbon dioxide produced being measured separately at each stage. The reducing temperature (210°c), initial pressure of carbon monoxide (300 to 400 mm) and duration of heating were, in general the same throughout all stages, the total time of heating being about 60 hours. In all 20·2 cm^3 at s.t.p. of carbon dioxide were collected, corresponding to the removal of 10·1 cm^3 of oxygen from the sample; the last and smallest individual quantity of gas collected was only 0·064 cm^3. The total quantity of carbon monoxide taken up agreed with the quantity of carbon dioxide collected to within 4%. For the largest individual measured quantities, where the measuring accuracy was greatest, agreement was obtained with 1%. During the initial outgassing treatment at 250°c a red deposit was seen to form on the cool portion of the adsorption cell directly above the heater. This was presumed to be selenium, but whether this was produced by dissociation of the lead salt or, on the other hand, by the evaporation of excess selenium from the sample was not clear. At the time of performing the experiments it was not possible conveniently to measure the amount of selenium removed, but this, however, appeared small compared with the quantity of powder in the cell.

§ 5. Absorption of Oxygen

Figure 2 shows graphically the results obtained for the oxidation of the lead selenide sample for temperatures in the range 19° to 280°c. Because the experiments were carried out in a constant volume system the oxygen pressure fell to some extent as oxidation proceeded. This decrease, however, was in general

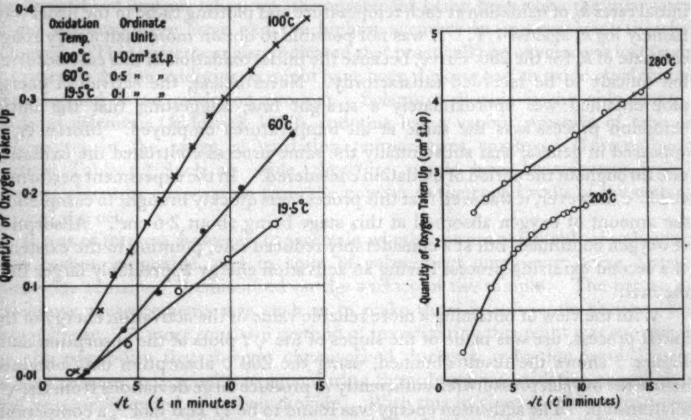


Figure 2. Oxidation of lead selenide at various temperatures.

kept small by using the burette bulbs to provide a large system volume compatible with reasonable accuracy in absorption measurements.

The significance of the quantities of oxygen taken up at the various temperatures can be realized approximately by a consideration of the number of lead or selenium ions lying in the surface of the lead selenide sample. The lattice spacing in a lead selenide crystal is 3·07 Å, and since the exposed surfaces of the particles were most likely to be cubic faces of the crystal, this then gives for the density of lead or selenium ions in the surface the approximate value $5{\cdot}5 \times 10^{14}$ per cm^2, or $3{\cdot}6 \times 10^{18}$ for the total number of each ion in the whole surface of area $6{\cdot}5 \times 10^3$ cm^2 as measured by argon adsorption. Therefore, the quantity of oxygen required for a monolayer in this case was of the order of 0·1 cm^3 at S.T.P. This quantity of oxygen uptake, it will be noted, was exceeded in all the experiments carried out above room temperature, about 4·5 cm^3 of gas being taken up in 5 hours at 280°C, showing that the reaction taking place was not confined to the surface.

It appeared in fact that the reaction proceeded by the diffusion of oxygen into the body of the crystallites. Anderson *et al.* (1955) have shown that diffusion into spherical particles follows the simple law :

$$Q \propto \sqrt{(Dt)},$$

where Q is the quantity of gas which has diffused in time t and D is the coefficient of diffusion at the temperature of the experiment, provided Q is much smaller than that corresponding to equilibrium. Figure 2 has been plotted on a $\sqrt{t}$ scale to show that the experimental data appear to conform to this law, except in the 200°C case when the oxygen pressure decreased by a large amount during the experiment. An attempt to represent the data by a logarithmic law ($Q \propto \log t$) did not give such a satisfactory result.

§ 6. THE ACTIVATION ENERGY OF THE OXIDATION PROCESS

The activation energy of the process was considered first by estimating the initial rates k_0 of oxidation at each temperature and plotting these in the usual way, namely $\log k_0$ against $1/T$. It was not possible to obtain more than a very rough estimate of k_0 for the 280° curve, because the initial oxidation in this case occurred too quickly to be recorded satisfactorily. Nevertheless, the activation energy plot obtained was approximately a straight line, suggesting that the initial oxidation process was the same at all temperatures employed. Moreover, it appeared in general that substantially the same process controlled the oxidation rate throughout the period of oxidation considered. In the experiment performed at 280°C, however, it was seen that this process was quickly brought to completion, the amount of oxygen absorbed at this stage being about 2·6 cm^3. Absorption of oxygen continued, but at a considerably reduced rate, pointing to the existence of a second oxidizing process having an activation energy appreciably larger than the first.

With the view of obtaining a more reliable value of the activation energy of the initial process, use was made of the slopes of the $\sqrt{t}$ plots of the absorption data. Figure 3 shows the result obtained, using the 200°C absorption data obtained before the oxygen pressure fell sufficiently to produce large deviations from the $\sqrt{t}$ relationship. The activation energy was found to be 17 kcal mol^{-1}, a comparable value being found from the consideration of k_0 values.

Now before discussing the possible mechanisms of the oxidation observed, it is necessary to consider the state of the lead selenide powder during the oxidation. It has already been pointed out that a visible though small amount of selenium was removed from the powder by the initial outgassing treatment. The possible further effects of heating in carbon monoxide on the composition of the powder and the state of the surface are discussed below.

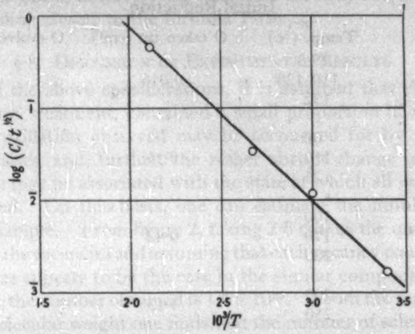


Figure 3. Activation energy curve for the initial diffusion process.

§ 7. THE SUITABILITY OF CARBON MONOXIDE AS REDUCING AGENT

7.1. *Presence of Chemisorbed Carbon Monoxide*

It has been noted above that during the initial reduction of the powder the quantity of carbon dioxide produced was in fairly good agreement with the corresponding quantity of carbon monoxide reacting, indicating that there was no appreciable side reaction taking place. In support of this, as shown in the table, it was found that the quantity of carbon dioxide produced agreed well with the quantity of oxygen taken up, the agreement being best when the quantities of gas involved were large and the highest accuracy in measurement thereby achieved. This agreement also indicated that practically no oxygen was lost from the reaction during oxidation as might have been the case had an oxide of selenium been formed as an oxidation product. Selenium dioxide is the only certain oxide of selenium (Sidgwick 1950), and this has a vapour pressure of tens of millimetres in the range of oxidation temperatures employed (International Critical Tables). Such an oxide, therefore, would have become transferred to some extent by evaporation from the powder to the cooler walls of the system beyond the cell.

It was of interest, however, to ascertain whether or not as a result of reduction with carbon monoxide, and in spite of subsequent outgassing, some carbon monoxide remained chemisorbed on the surface of the sample. The nature of the above measurements was such as to make them insufficiently accurate for this purpose. A more sensitive method of investigating this point was suggested by the possibility that the gas chemisorbed during a reduction would react, at least to some extent, with oxygen during the subsequent oxidation experiment and become liberated as carbon dioxide. With this in view the gases remaining at the end of each oxidation were pumped slowly through the cooled capillary trap, and the quantity of condensable gas trapped was subsequently measured. By this means quantities of gas as small as 8×10^{-4} cm^3 at S.T.P. could be detected. With the exception of the 280°C oxidation the quantity of gas collected in this way was never greater than this order of magnitude.

Summary of Oxidation and Reduction Experiments carried out after Initial Reduction

	Temp. (°C)	O taken up (cm^3)	O evolved as CO$_2$ (cm^3)
Oxidation	19·5		Zero
Reduction	110–120	0·016	0·015
Outgassing	132		
Oxidation	60		Zero
Reduction	110	0·16	0·13
Outgassing	126		
Oxidation	100		Zero
Reduction	170	0·42	0·33
Outgassing	200		
Oxidation	200	4·13	Zero
Reduction	250		4·10
Reduction	330		0·03
Outgassing	350		
Oxidation	280	4·68	0·005
Reduction	330		4·70
Outgassing	200		

7.2. *The Possible Formation of Carbon Oxyselenide during Reduction*

It has been assumed throughout in the above that carbon dioxide was the only product of heating the lead selenide sample in carbon monoxide. The present measurements, however, do not provide sufficient evidence to preclude entirely the possibility of carbon oxyselenide formation (COSe), this gas being capable of being collected in a liquid air trap in the same way as carbon dioxide. On the one hand, it appears unlikely that an appreciable quantity of carbon oxyselenide was formed because even during the preparation of this compound, by passing carbon monoxide over free selenium at red heat, the reaction does not proceed rapidly, and, furthermore, the compound though stable in the cold is easily decomposed to CO + Se when heated (Sidgwick 1950). On the other hand, however, the total quantity of condensable gas collected during the initial reduction of the sample, if assumed to be entirely carbon dioxide, points to a rather large quantity of oxygen present in the sample after grinding when compared with the quantity of oxygen taken up by the reduced powder after 3 hours oxidation at 280°C, the respective figures being about 10 cm^3 and 4 cm^3. It seems necessary, therefore, when considering the oxidation measurements to keep in mind the possibility that, though the original crystal was p-type, the material being oxidized contained selenium vacancies.

Another point is of interest here, however. After the initial reduction of the powder at temperatures of about 210°C, yielding about 20 cm^3 of condensable gas, the later reduction of 330°C, as shown in the table, yielded only 0·06 cm^3 of condensable gas. Taking the view that selenium was being removed from the sample during the initial reduction, this result indicates that at the end of this reduction there had been formed a stable lead–selenium phase from which further selenium could not easily be removed. The stability of this phase is suggested also by the observation that oxygen could be introduced into and removed from the material in an apparently accurately reversible manner. The phase diagram (Jänecke 1949) for PbSe, however, indicates that the only stable phase obtained is that corresponding closely to the formula $PbSe_{1\cdot 00}$.

§ 8. DISCUSSION OF EXPERIMENTAL RESULTS

If, in view of the above considerations, it is assumed that the lead selenide sample, after initial treatment, contained a small proportion of vacant selenium sites, the initial oxidation observed may be accounted for by the diffusion of oxygen to these sites, and, further, the rather abrupt change in oxidation rate obtained at 280°c may be associated with the state at which all or nearly all these sites became filled. On this basis, one can estimate the number of selenium vacancies in the sample. From figure 2, taking 2·6 cm^3 as the quantity of oxygen required to fill all the vacancies and assuming that each vacancy could accommodate one oxygen ion, as appears to be the case in the similar compound lead sulphide (Anderson 1948), the number obtained is $1\cdot 4 \times 10^{20}$. From the mass of the sample (3·8 g) and its molecular weight one finds that the number of selenium ions in the stoichiometric sample was about $8\cdot 0 \times 10^{21}$, so that the proportion of vacant selenium sites was about 1/57. This would give for the composition of the stable phase referred to above the formula $PbSe_{0\cdot 98}$, differing only slightly from stoichiometry.

Turning to the experiments on lead selenide layers (Jones 1957), these indicated that the layer material also, being n-type, was initially deficient in selenium. The initial rapid fall in conductivity observed on exposing a layer to oxygen was thus accounted for by the filling-up of selenium vacancies until a quasi-stoichiometric state was reached at the lowest conductivity value. The conductivity, however, did not remain constant at this value, but now began to increase slowly, pointing to further absorption of oxygen at a reduced rate. The layer studies, therefore, support qualitatively the above interpretation of the powder measurements. It must be pointed out, however, that on this basis one would have expected the activation energy deduced from the powder measurements to have been smaller than that obtained from the layer measurements, since these energies refer respectively to conditions obtaining before and after the filling of selenium vacancies. The layer experiment value, in fact, was only 14 kcal mol^{-1} which is somewhat smaller than the value deduced above. It is not clear, however, that the values obtained were sufficiently accurate to enable them to be compared in this way.

As has been implied above, there was no evidence in the present powder measurements of a rapid surface adsorption of oxygen as the initial stage in the oxidation of the freshly reduced sample. The results instead indicated a continuous diffusion of oxygen into the sample beginning when the quantity of oxygen taken up was at least an order of magnitude smaller than that required to form a monolayer of oxide, this being shown most clearly in the case of slow oxidation obtained at room temperature. It is of interest now to consider this in relation to the mechanism of photoconductivity in evaporated lead selenide films sensitized by oxygen. Two possible mechanisms have been suggested (Smith 1951). One postulates for n-type layers the existence of inter-crystallite potential barriers in the film, produced by the adsorption of oxygen on the surfaces of the crystallites. On this view the predominant effect of irradiation is to eject electrons from the oxygen surface states thereby lowering the barrier heights and producing an increased film current by increasing the carrier mobility. The other mechanism proposed is based on the induced transitions in the crystallites themselves. In this case oxygen absorbed is supposed to act as a bulk acceptor impurity in that the barrier formed is low and the Fermi level is altered throughout each crystallite. The effect of irradiation in this case is to increase the number of carriers rather than their mobility. The present measurements, indicating bulk oxidation rather than oxidation confined to the surface region of the powder particles, therefore support the 'numbers' view of photoconductivity rather than the barrier theory. The observations on evaporated layers also were in accordance with this conclusion, since the effect of exposing a vacuum-baked layer to oxygen was to produce throughout the exposure a gradual resistance change which showed no marked discontinuity which could be attributed to initial rapid chemisorption.

§ 9. CONCLUSIONS

The above measurements indicate that in the oxidation of a lead selenide sample prepared in the manner described, the mechanism of the oxygen absorption is predominantly the diffusion of oxygen ions to selenium vacancies in the lattice. At sufficiently high temperatures, however, a second absorption process is revealed. This, like the first, yields a parabolic oxidation law and may, therefore, also be a simple diffusion process, the rate of diffusion being now decreased owing to the previous filling up of selenium vacancies. This view of the oxidation process conforms qualitatively with what has been observed in lead selenide layer studies, and supports the 'numbers' theory of photoconductivity in layers sensitized by oxygen rather than that depending on intercrystalline potential barriers.

ACKNOWLEDGMENTS

The author is indebted to colleagues in this laboratory and to members of the Chemistry Division, Atomic Energy Research Establishment, Harwell, for helpful advice and discussion during the course of this work. Acknowledgment is made to the Controller, H. M. Stationery Office, for permission to publish this paper.

REFERENCES

ANDERSON, J. S., 1948, *Disc. Faraday Soc.*, **4**, 163.
ANDERSON, J. S., ROBERTS, L. E. J., and HARPER, E. A., 1955, *J. Chem. Soc.*, 3946.
BRUNAUER, S., 1944, *The Physical Adsorption of Gases and Vapours* (Oxford : Clarendon Press).
BRUNAUER, S., EMMETT, P. H., and TELLER, E., 1938, *J. Amer. Chem. Soc.*, **60**, 309.
INTERNATIONAL CRITICAL TABLES, 1928, **3**, 207.
JÄNECKE, E., 1949, *Handbuch aller Legierungen* (Heidelberg : Universitätsverlag), p. 195.
JONES, R. H., 1957, *Proc. Phys. Soc.* B, **70**, 704.
SIDGWICK, N. V., 1950, *The Chemical Elements and their Compounds* (Oxford : Clarendon Press).
SMITH, R. A., 1951, *Semiconducting Materials*, Ed. H. K. Henisch (London : Butterworths Scientific Publications), p. 198.

The Dynamic Stress-Optic Coefficient of Perspex

By D. F. MARSHALL
Imperial Chemical Industries Limited, Akers Research Laboratories, Welwyn, Herts.

Communicated by J. J. Benbow; MS. received April 4th 1957, *and in revised form August 20th* 1957.

Abstract. An apparatus based upon the Davies and Kolsky modifications of the Hopkinson pressure bar, and including a plane polariscope, was used to measure the dynamic stress-optic coefficient of a bar of Perspex whilst the material was being subjected to a stress cycle of about 50 microseconds duration. A value of 11 Brewsters was obtained for the coefficient of the particular sample of Perspex used, compared with a value of 23·4 Brewsters obtained by subjecting the same specimen to a static load.

---

§ 1. INTRODUCTION

MODERN methods of high speed photography using multiple sparks (Cranz and Schardin 1929; see also Christie 1952) or 'streak' (Tuzi and Nisida 1936; see also Frocht and Flynn 1956) techniques have widened considerably the scope of dynamic photoelastic investigations. It has become desirable to know the value of the stress-optic coefficients of materials employed in the investigation of such high speed phenomena as the propagation of shock waves and associated cracks. Previous work (Senior and Wells 1946) has shown the possibility that for Xylonite the dynamic stress-optic coefficient differs by as much as 38% from the static value, whilst Frocht and Flynn (private communication to Dr. H. Kolsky) have observed a similar decrease for Castolite. There is of course no reason for assuming that the stress-optic coefficient is independent of rate of straining. This paper describes the investigation into the value of the dynamics tress-optic coefficient of unplasticized Perspex—a material frequently used in photoelastic experiments, particularly in this laboratory.

The method described here has the advantage that, unlike that used by Senior and Wells, no high speed photographic techniques are necessary, other than that involved in photographing a cathode-ray trace. Furthermore, the analysis of fringes is both simple and more accurate by photomultiplier than by photographic reproduction. In its present form, the apparatus is limited to the measurement of the effects caused by plane longitudinal waves only, however as such it can be used as a reasonably simple standard test apparatus for a number of different materials.

§ 2. DESIGN OF APPARATUS

The apparatus, the arrangement of which is shown in figure 1, consisted of two principal sections. One for the application of a pressure pulse to the specimen and for the electrical and mechanical measurement of the pulse; the other section for the optical measurement of the photoelastic effects caused by the passage of the pulse through the specimen.

A pressure pulse of approximately 7×10^7 dyn cm^{-2} maximum pressure and about 50 microseconds duration was produced in a 1·5 m long by 2·54 cm diameter cylindrical Hopkinson pressure bar made of tool steel, by the detonation of a 1 g pellet of Tetryl in contact with a mild steel anvil wrung to one end of the bar with a thin layer of grease. The pulse was transmitted to the test bar, which was a 58 cm long bar of Perspex of 1·27 cm square cross section, by wringing the test bar to the end of the pressure bar remote from the explosion. The test bar thus formed an extension to the pressure bar. The ends of the two bars in contact were turned flat and alignment was ensured by using a mild steel 'adaptor' to hold the bars together. This precaution was necessary so that no distortion of the pressure pulse, which was assumed plane, should occur at the interface.

The free or measuring end of the test bar was faced and several centimetres of the surface of the bar at that end were coated with a thin layer of graphite to form a conducting surface over the sides and end face. This was done in order to measure the pressure pulse—the displacement of the free end of the bar being detected with a parallel plate condenser microphone of the type used by both Davies and Kolsky.

For optical measurements a plane polariscope was used with a narrow, approximately parallel beam of monochromatic light obtained from a high pressure mercury arc lamp and a green filter transmitting 5461 Å wavelength. A condenser lens, shown in figure 1 as C, was used to focus the image of a circular aperture A placed over the source on a plane just outside the focal plane of a microscope objective O. The lens O thus formed a diminished image of the aperture, with the light beam converging at an angle of approximately 4·5°. The test bar was hung with its two polished sides perpendicular to the axis of the beam and the image of the source at its centre. In this position the beam converged from 1 mm diameter at each face of the bar to 0·5 mm diameter in the centre of the bar. An RCA 931A photo-multiplier operating at 75 volts per stage was used to detect changes in intensity of the light beam caused by the passage of a stress pulse along the bar. The principal directions of the polarizer and analyser were set at an angle of 45° to the axis of the test bar. Assuming the pressure pulse to be plane this ensured that the direction of the principal stress also made an angle of 45° with the polarizer and analyser.

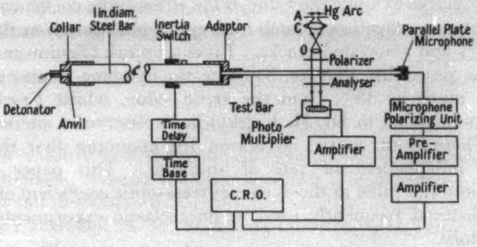


Figure 1. General arrangement of apparatus.

To eliminate as far as possible any slight change in the pulse in its passage between optical and electrical measurements the light beam was made to intercept the test bar near the measuring end of the bar. It was necessary, however, to allow the tail of the pressure pulse to 'clear' the light beam before the head of the pulse—reflected from the free end as a pulse of tension—reappeared in the beam. The distance between the light beam and the free end of the test bar was also sufficient to enable the velocity of the pulse to be measured by observing the time between the initial appearance of the pulse in the beam and its reappearance after reflection.

Amplified recordings of the outputs of the condenser microphone and the photo-multiplier were displayed on a double beam cathode-ray oscilloscope and photographed. A standard timing trace, and a trace for determining the amplification factor of the condenser microphone circuits were displayed and photographed immediately after each experiment. The oscilloscope was triggered by a conventional inertia switch on the pressure bar with suitable delays incorporated to ensure correct synchronization.

§ 3. THEORY OF METHOD

3.1. *Pressure Bar*

Davies (1948) has shown that for a plane pressure pulse propagated in a cylindrical bar, provided the elastic limit is not exceeded and that the duration of the pulse is long compared with the time taken to travel a distance equal to the radius of the bar, the following relation holds good

$$p=\tfrac{1}{2}\rho c\frac{dx}{dt} \qquad \ldots\ldots(1)$$

where p is the pressure at time t, ρ the density of the bar, c the velocity of propagation of the pulse and x the displacement of the end of the bar at time t.

Davies has also shown that for the parallel plate condenser microphone, provided the displacement of the moving plate is small compared with the separation between the plates then

$$v=\frac{VC_1x}{CD} \qquad \ldots\ldots(2)$$

where v is the change in voltage, x the displacement of the moving plate, V the polarizing voltage, C_1 the capacity of the microphone, C the capacity of the microphone plus all the shunt capacities and D the initial separation between plates.

Differentiating (2) we have

$$\frac{dv}{dt}=\frac{VC_1}{CD}\frac{dx}{dt}. \qquad \ldots\ldots(3)$$

From equation (1)

$$p=\tfrac{1}{2}\rho c\frac{CD}{VC_1}\frac{dv}{dt}. \qquad \ldots\ldots(4)$$

By integrating equation (1) it can be seen that the maximum displacement $x_{\max}$ is given by

$$x_{\max}=\frac{2}{\rho c}\int p\,dt. \qquad \ldots\ldots(5)$$

Now if M is the momentum associated with the pulse, and A is the cross-sectional area of the bar

$$M = A\int p\,dt. \qquad \ldots\ldots(6)$$

Thus from (5) and (6)

$$x_{\max} = \frac{2M}{A\rho c}. \qquad \ldots\ldots(7)$$

Also,

$$M = \frac{2\pi mX}{T} \quad \text{(Davies 1948).} \qquad \ldots\ldots(8)$$

Hence from (7)

$$x_{\max} = \frac{4\pi mX}{A\rho cT} \qquad \ldots\ldots(9)$$

where m is the mass of the bar, X the maximum amplitude of swing and T the ballistic period of swing of the bar.

3.2. *Polariscope*

For a plane polariscope of dark field, the amplitude a of the wave emerging from the analyser is given by

$$a = A\sin 2\alpha \sin(\pi r/\lambda) \qquad \ldots\ldots(10)$$

where A is the amplitude of the incident light, α the angle between the principal axes of stress and the plane of polarization, λ the wavelength of light and r the relative retardation of the ordinary and extraordinary rays through the specimen.

Thus, if I is the intensity of light incident upon the photo-multiplier, and α remains constant at 45°, then

$$I = B\sin^2(\pi r/\lambda) \qquad \ldots\ldots(11)$$

where B is a constant.

Maximum values of I occur for values of $r = \lambda/2$, $3\lambda/2$ etc. . . . Minimum values of I occur for values of $r = 0$, λ, 2λ, etc. . . . Intermediate values of r for any recorded intensity between $I_{\max}$ and $I_{\min}$ may be obtained from a graph of the function $\sin^2(\pi r/\lambda)$ against r provided r is known for $I_{\max}$ and $I_{\min}$.

The relation between relative retardation and stress is

$$r = Kpd \qquad \ldots\ldots(12)$$

where K is the stress-optic coefficient, and d is the length of light path in specimen. If r is measured in Angström units, p in megadynes cm^{-2} and d in millimetres, K is given in Brewsters and has the dimensions of the reciprocal of force.

§ 4. METHOD

Morse (1948) has shown that the dispersion of a plane longitudinal wave in a rod of square section is indistinguishable from that in a rod of circular cross section. Equation (1) has therefore been applied directly to these experiments.

The constants of the apparatus were measured before and after the series of experiments and were not found to have varied. The shunt capacity of approximately 0·005 microfarad was measured with a Sullivan and Griffiths capacitance bridge Type AC1101 with accuracy to 0·2%. V was measured by valve voltmeter and found to be 340 v ± 2%. The capacity C_1 of the microphone was plotted against $1/D$ for values of D between 0·051 cm and 0·114 cm and the value used taken from this graph. Gaps of 0·089 cm and 0·076 cm only were used in the experiments. The value of y—the distance between the beam of light and the free end of the test bar—was measured with the bars clamped and polaroids temporarily uncrossed, the range of values being 16 to 18 cm. Feeler gauges were used to set D after this operation with the bar unclamped and a correction to y was applied to allow for movement of the bars during the setting of D. The height of the test bar was adjusted by eye to allow the beam of light to pass through the centre of the bar, and only occasional correction was necessary to maintain this condition.

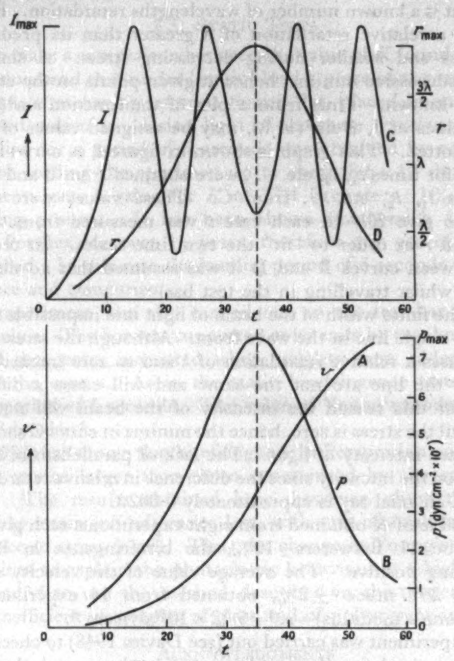


Figure 2. (t is in μsec.)

To eliminate all light other than the monochromatic beam, the experiment was carried out in a darkened room, the flash from the explosion being of no consequence, since a delay of some 500 microseconds occurred between the explosion and triggering of the recording apparatus.

Immediately after each experiment a standard 100 kc/s time trace and a 1·5 Mc/s 10 mv signal were displayed upon the cathode-ray oscilloscope, the latter signal being connected directly across the output terminals of the condenser microphone and giving a measure of the amplification of the associated circuit.

Direct plots of the amplified outputs of the condenser microphone and photo-multiplier were made by projecting the photographic records on to graph paper. In figure 2 these curves, marked A and C respectively, are shown diagrammatically together with the curves obtained from them by calculation. Numerical differentiation with respect to t of curve A, which is a graph of v against t gave curve B representing p against t (4). From curve C, which is a graph of I against t, and using a graph of $\sin^2(\pi r/\lambda)$ against r the curve D was obtained, showing the variation of r with t.

Since the stress and birefringence return to zero it is possible to assign a value of r to each maximum and all except one minimum in the curve of I against t. The central minimum may not be given a value of r by inspection since it corresponds to the retardation at maximum stress and not necessarily to a known fringe order that is a known number of wavelengths retardation. Each maximum corresponds to a relative retardation of λ greater than its predecessor during increasing stress and smaller during decreasing stress. A similar argument applies for the successive minima, hence at given points on the curve I against t values of r are known. Thus from a plot of the function $\sin^2(\pi r/\lambda)$ against r intermediate values of I, $B\sin^2(\pi r/\lambda)$, may be assigned values of r and a graph of r against t plotted. This graph is shown in figure 2 as curve D.

Values of p for times t_1, t_2, etc. . . . were obtained from B and values of r for the same times t_1, t_2, etc. . . . from C. These values were substituted in equation (12) to give K. In each case t was measured from the position of maximum p and r in order to 'fit' the two time scales. In order to make a comparison between curves B and D it was assumed that no distortion of the pulse occurred whilst travelling in the test bar.

Because of the finite width of the beam of light it is impossible to measure the stress along any single line in the wave front. Although the stress along the axis of the beam causes a relative retardation of λ and so zero intensity of light, the stresses around the line are not the same and will cause a different relative retardation. For this reason the intensity of the beam will not return to its original zero until the stress is zero, hence the minima in curve C show an apparent change in the zero intensity of light. The lack of parallelism of the light has a negligible effect on the intensity since the difference in relative retardation between an edge ray and an axial ray is approximately $0{\cdot}002\,\lambda$.

The mean value of K obtained from eight experiments each giving ten values was found to be 11 Brewsters ± 10%, the birefringence in Perspex under compression being positive. The average value of the velocity c used for the calculation was $2077\,\mathrm{m\,sec^{-1}} \pm 2\%$, obtained from 14 experiments, giving a value for E (Young's modulus) $= \rho c^2 = 5{\cdot}12 \times 10^{10}\,\mathrm{dyn\,cm^{-2}}$.

A ballistic experiment was carried out (see Davies 1948) to check the accuracy of the calibration used for the condenser microphone and the amplification factor signal. Equation (7) shows that the maximum displacement $x_{\max}$ of the end of the pressure bar remote from the explosion may be measured by determining the momentum associated with the pressure pulse. This was done by firing a detonator and measuring the maximum amplitude X and period of oscillation T of the bars, at the same time recording the output of the condenser microphone. Ideally the theory deals only with a single bar; however, it was essential to have the test bar in place in order to make the electrical measurements since no two charges could be guaranteed to give identical effects and it was necessary to measure the displacement electrically *and* ballistically. Since the mass of the test bar was only 1% of the total mass its effect was neglected and the calculations of momentum applied purely to the steel bar. The value for $x_{\max}$

obtained from momentum considerations, by substitution in equation (9), was found to be 0·025 cm ± 5%. The displacement of the two ends of the bars in contact during compression is the same, hence the maximum displacement of the pressure end of the test bar is equal to $x_{\max}$. Since the displacement of the free end of the bar depends only upon the momentum associated with the pulse it also will be $x_{\max}$. From the electrical measurements, and the values of the constants previously measured, the displacement of the free end of the test bar was also found to be 0·025 cm ± 5%. Thus the constants of the apparatus were verified.

§ 5. STATIC TESTS

In order to make a comparison between the value of K found under the above loading conditions and that under 'static' load a piece of the test bar 4 cm long was cut off and, after having its ends turned flat, placed in an hydraulic press between crossed polaroids. Using light of the same wavelength a value of K for loads applied for a period of 5 seconds was obtained by counting the dark fringes seen on loading the specimen up to about $8{\cdot}6 \times 10^7$ dyn cm^{-2}. Several such experiments gave a value for K of 23·4 Brewsters ± 1%, again the birefringence being positive. It has been shown (Lethersich 1950) that creep continues in Perspex for over $1\frac{1}{2}$ years, therefore any use of the word static must be qualified. In this case the loading period of 5 seconds was taken purely for convenience and comparison.

A static experiment was also carried out on the test bar to determine its Young's modulus. The bar was supported at its ends by two knife edges and the depression of the centre measured when weights were added to a scale pan suspended at the centre. As in the static birefringence tests the loads were applied for a period of 5 seconds. The value of Young's modulus was found to be $3{\cdot}05 \times 10^{10}$ dyn cm^{-2}.

Kolsky (1952) has shown that birefringence in Perspex is a uni-valued function of strain for strains larger than those experienced by the specimen in these experiments. The results obtained here, however, show that the ratio of birefringence to strain (rE/pd) varies less, with rate of straining, than the ratio of birefringence to stress (r/pd). Thus the change in Young's modulus appears to be insufficient to explain the photoelastic behaviour at a high rate of loading.

It is intended to use the method described in this paper to measure the stress-optic coefficient of samples of glass and of polystyrene in the near future.

ACKNOWLEDGMENTS

The author wishes to thank Mr. D. G. Christie and Dr. F. C. Roesler for helpful discussions during this work.

REFERENCES

CHRISTIE, D. G., 1952, *J. Soc. Glass Tech.*, **36**, 74.

CRANZ, C., and SCHARDIN, H., 1929, *Z. Phys.*, **56**, 147.

DAVIES, R. M., 1948, *Phil. Trans. Roy. Soc.* A, **240**, 375.

FROCHT, M. M., and FLYNN, P. D., 1956, *J. Appl. Mech.*, **23**, 116.

KOLSKY, H., 1949, *Proc. Phys. Soc.* B, **62**, 676; 1952, *J. Soc. Glass Tech.*, **36**, 56.

LETHERSICH, W., 1950, *Brit. J. Appl. Phys.*, **1**, 294.

MORSE, R. W., 1948, *J .Acoust. Soc. Amer.*, **20**, 833.

SENIOR, D. A., and WELLS, A. A., 1946, *Phil. Mag.*, **37**, 463.

TUZI, Z. H., and NISIDA, M., 1936, *Phil. Mag.*, **21**, 448.

A Determination of the Surface Tension of Liquid Tungsten by the Drop-Weight Method

By A. CALVERLEY

Services Electronics Research Laboratories, Baldock, Herts.

M.S. received 28*th May* 1957, *and in final form* 6*th September* 1957

Abstract. The production of molten drops by bombarding the lower end of a vertically suspended rod with electrons has enabled the drop-weight method to be applied to the measurement of the surface tension γ of liquid tungsten. The mean value obtained is 2300 dyn cm^{-1} at the melting temperature of 3380°c in a vacuum of approximately 10^{-4} mb.

---

§ 1. Introduction

Experimental difficulties have limited the determination of liquid surface tension to about half the metallic elements and the highest melting point metal so far reported (Yelutin and Maurakh 1956) is titanium (melting point 1820°c, $\gamma = 1510$ dyn cm^{-1}). Surface tension values are of interest both to the technologist and to the theoretician, and, although values for twenty seven metals have been estimated by Taylor (1954) with a probable accuracy to $\pm 10\%$ in some cases, experimental verification of the value of γ for any liquid refractory metal has not been attempted. For tungsten, Taylor's estimated value is 2680 dyn cm^{-1}.

Of the methods available, the drop weight method is one of the most accurate, but the higher melting point metals have not been amenable to this method (or any other) because of the difficulty in raising them to liquid temperatures in a controllable manner.

Experience with electron bombardment floating zone refining (Davis, Calverley and Lever 1956) showed that liquid drops could be readily formed on the end of a vertically suspended refractory metal rod and hence suggested its application to surface tension measurements by the drop-weight method.

§ 2. Apparatus

The essential part of the apparatus is shown in figure 1 and is generally as described elsewhere (Calverley, Davis and Lever 1957). Two additions are a molybdenum cup X and a nickel shield Y which can be rotated from outside the vacuum wall to expose the cup to the drop. The specimen S is held in a V groove in the block G so that its lower end is opposite the tungsten wire cathode B. The latter is sandwiched between molybdenum focusing plates F which are at the same potential as the cathode and reduce the power necessary to produce a molten drop. To assist further in this reduction, a molybdenum mesh grid W is hung from the top focusing plate. The whole electrode system is suspended from a steel top plate A and is mounted in a vacuum chamber.

The specimen is held at earth potential and the cathode at 2 or 3 kv negative, so that by raising the temperature of the cathode, electron bombardment of the specimen tip takes place and it can easily be made molten.

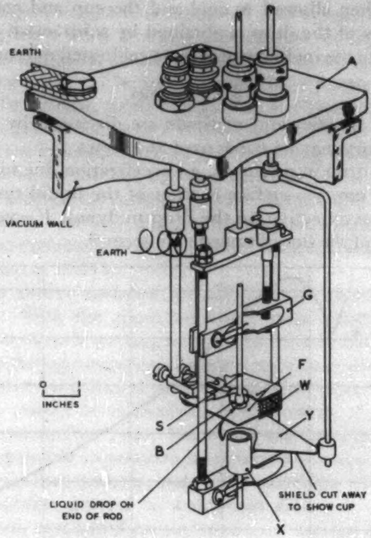


Figure 1. The electrode system.

§ 3. METHOD

The specimen, in this case a tungsten rod, is mounted in the apparatus and the molybdenum cup is weighed and placed in position under it. The nickel shield is rotated to cover the cup and the electrode system placed in the vacuum chamber which is then evacuated. The cathode heater current is switched on to a preset value and the bombarding voltage increased until the tip of the rod becomes molten.

If the rod has not been previously melted *in vacuo*, considerable outgassing occurs when it becomes molten and some material is lost from it in the form of a shower of sparks. It is mainly for this reason that the loss in mass of the rod cannot be used for determining the mass of the drop. The nickel shield was incorporated to ensure that none of the tungsten thrown out during outgassing entered the cup.

As the size of the drop increases, the rod shortens and the specimen frame has to be lowered to keep the drop in the bombarding region. For smaller diameter rods this shortening amounts to several centimetres and in these cases the tungsten is allowed to cool, the system is let down to air and the cup raised to restore the distance between it and the drop to its former value. The system is then reevacuated and the heating continued.

When the drop is becoming large enough to fall (this is judged by experience), the nickel shield is rotated to clear the cup. To avoid shaking the drop off during this operation, the power is switched off to allow the drop to solidify. The heating is then recommenced and the rotary backing pump switched off, so that when the drop falls vibration of the system has been reduced to a minimum.

The system is then allowed to cool and the cup and contents removed and weighed. The mass of the drop is obtained by subtraction and the diameter of rod measured close to the molten end using a calibrated micrometer.

§ 4. THEORY

Surface tension theories and methods are discussed by Harkins (1949) and the drop weight theory may be condensed as follows :

Let m = mass of drop in grammes, g = acceleration due to gravity in cm sec^{-2}, r = radius of rod in cm, γ = surface tension of the liquid tungsten in dyn cm^{-1}, W = force due to gravity acting on the drop in dynes, V = volume of the drop in cm^3, ρ = the density of the liquid tungsten in g cm^{-3}.

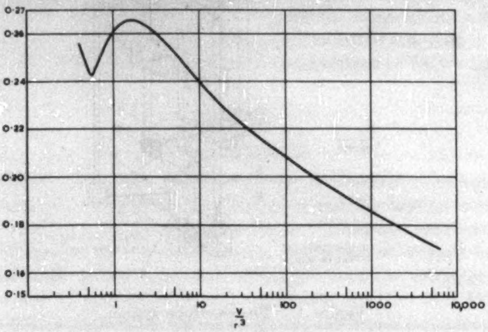


Figure 2. $F=F(V/r^3)$ (Harkins and Brown 1919).

If the drop hanging on the rod were cylindrical and of radius r, then the maximum weight of drop which could be supported would be equal to the weight of liquid which could be supported in a capillary tube of internal radius r. For in both cases, the force of surface tension acts on a line $2\pi r$ long and equals $2\pi r\gamma$

i.e. $$W=mg=2\pi r\gamma. \qquad \ldots\ldots(1)$$

Both observation and theory indicate that only a fraction of the drop falls, becoming larger as the tip is made smaller. The drop of weight $2\pi r\gamma$ is called the 'ideal drop'. The fraction of it which falls has been determined by Harkins and Brown (1919) and is a function of $(r/V^{1/3})$.

Thus

$$W=f(r/V^{1/3})2\pi r\gamma \qquad \ldots\ldots(2)$$

or since $W=mg$

$$\gamma=\frac{mg/2\pi r}{f(r/V^{1/3})} \qquad \ldots\ldots(3)$$

or

$$\gamma=F\,mg/r \qquad \ldots\ldots(4)$$

where F is a function of V/r^3 (or $m/\rho r^3$).

The values of F corresponding to given values of V/r^3 are given in *International Critical Tables* Vol. IV and are plotted in figure 2.

§ 5. ERRORS

The value of r used in the calculation is that value determined by micrometer when the rod is cold to which is added an estimated 2·5% to allow for expansion.

The density of liquid tungsten is unknown and a value of 10% change in volume in changing from solid at room temperature to liquid at the melting temperature has been estimated (i.e. $3 \times 2{\cdot}5\%$, $+2{\cdot}5\%$ on melting). A value of $\rho_{\text{liquid}} = 17{\cdot}6\ \text{g cm}^{-3}$ was used. The variation of $F(V/r^3)$ with error in estimation of density is small. In one case, recalculation of $F(V/r^3)$ assuming a volume change of 8% instead of 10% gave a change in F of less than 0·25%.

The liquid tungsten is assumed to be at the melting temperature. Previous observations on electron bombarded molten zones have shown at least the surface of the liquid to be in continuous motion and this may prevent a steep temperature gradient being formed. In any case, accurate determination of the liquid temperature would be extremely difficult and the value of γ is therefore taken to be that for liquid tungsten at 3380°c.

The drop-weight theory assumes the solid–liquid interface to be perfectly flat and horizontal. With the apparatus used, these conditions could not be guaranteed and the variation in γ owing to this imperfection has not been estimated. Vibration of the drop during outgassing for previously unmelted specimens almost certainly leads to premature falling of the drop and a low value for γ, and here again the percentage error cannot be estimated. To eliminate this error, an attempt was made to use zone melted rods, but these are invariably of non-uniform diameter and it is impossible to measure accurately the diameter of the rod at the 'partition' point on a non-uniform rod. To overcome this problem, a zone melted rod was centreless ground to a uniform diameter. This is a difficult procedure and was only successful on one specimen; it did however obviate the outgassing. If time allowed, this should be the procedure in all experiments.

§ 6. RESULTS

The results are summarized in the table. The mean value of the 15 results is 2280 dyn cm^{-1}. The low value of 2110 dyn cm^{-1} obtained on two occasions is

Summary of Results

Nominal diameter (mm)	Condition of rod	Determination No.	γ (dyn cm^{-1})	Mean γ (dyn cm^{-1})
3·0	Commercial quality centreless ground	(i)	2220	2220
2·0	,, ,,	(i)	2290	2328
	,, ,,	(ii)	2320	
	,, ,,	(iii)	2380	
	,, ,,	(iv)	2320	
	,, ,,	(v)	2330	
1·5	,, ,,	(i)	2340	2285
	,, ,,	(ii)	2110	
	,, ,,	(iii)	2350	
	,, ,,	(iv)	2340	
1·0	,, ,,	(i)	2260	2223
	,, ,,	(ii)	2210	
	,, ,,	(iii)	2110	
	,, ,,	(iv)	2310	
1·5	Zone melted, centreless ground	(i)	2330	2330

almost certainly due to excessive outgassing and if these are ignored the mean value is 2310 dyn cm^{-1}.

The last value obtained, 2330 dyn cm^{-1}, was with a zone melted rod, centreless ground and the drop appeared to be free from any vibration.

The vacuum throughout the experiment was approximately 10^{-4} mb.

The mean values obtained with various size rods suggest there may be a variation of γ obtained with the diameter of the rod used, but the spread of results does not justify such a conclusion.

To assess the method before carrying out a series of experiments on tungsten, a single determination was made on a 6 mm diameter copper rod. The value obtained was 1150 dyn cm^{-1} at the melting temperature of 1083°c. This compared favourably with the value of 1140 dyn cm^{-1} measured by the maximum bubble pressure method in hydrogen and obtained from the *International Critical Tables*.

§ 7. CONCLUSIONS

The mean value obtained for the surface tension of liquid tungsten at its melting point and in a vacuum of 10^{-4} mb is 2300 dyn cm^{-1}. It is difficult to assign an overall error to the measurement, but $\pm 5\%$ would seem a fair estimate.

The method is applicable to most of the high melting point metals which are amenable to electron bombardment, i.e. the surface tension of such metals as rhenium, tantalum, molybdenum, iridium, rhodium and ruthenium can be readily determined providing they can be obtained as smooth rods of a few millimetres diameter. The method, as it stands, does not permit a determination of the variation of γ with temperature.

A point of merit for these determinations is that the molten surface, being in a vacuum, is free from contamination, for it is well known that surface contamination can considerably influence the value of γ.

The apparatus used, being an adaptation of an existing apparatus, is not ideal. For further work on the refractory metals, it would be more convenient to have the molybdenum cup at a fixed distance from the bombarding region with provision for the end of the specimen to be slowly and smoothly lowered into the bombarding region. The nickel shield should be mounted so that it too, could be moved without vibrating the apparatus.

With special attention to electrode configuration to obtain uniform heating of the tip, and by using centreless ground zone melted rods, the method could yield results of higher accuracy.

ACKNOWLEDGMENTS

The author would like to thank Mr. R. F. Lever for his interest and encouragement given in the early stages of the experiment.

Permission from the Admiralty to publish this paper is also acknowledged.

REFERENCES

CALVERLEY, A., DAVIS, M., and LEVER, R. F., 1957, *J. Sci. Instrum.*, **34**, 142.
DAVIS, M., CALVERLEY, A., and LEVER, R. F., 1956, *J. Appl. Phys.*, **27**, 195.
HARKINS, W. D., 1949, *Technique of Organic Chemistry* 2nd Edition, Vol. 1, part I (London : Interscience Publishers). Chap. IX.
HARKINS, W. D., and BROWN, F. E., 1919, *J. Amer. Chem. Soc.*, **41**, 499.
TAYLOR, J. W., 1954, *Metallurgia, Manchr*, **50**, 161.
YELYUTIN, V. P., and MAURAKH, M., 1956, *Investiya Akad. Nauk, SSSR, Otdelenie Teckhnicheskikh Nauk*, **4**, 129.

The Variation with Temperature of the Dynamic Elastic Moduli of an Organic Glass at Two Frequencies

By S. M. CRAWFORD†

Imperial Chemical Industries Limited, Akers Research Laboratories,
Welwyn, Herts.

MS. received 17*th June* 1957, *and in final form* 22*nd July* 1957

Abstract. Measurements have been made of the dynamic shear modulus and velocity of longitudinal waves in an organic glass at frequencies of 4 Mc/s and 20 Mc/s. It has been found that there is a relaxation in the real part of the shear modulus which occurs at a higher temperature at 20 Mc/s than at 4 Mc/s; at each of these temperatures, the relaxation time is of the same order as the period of the applied stress. The real part of the bulk modulus calculated from the results appears to be independent of frequency but shows a transition in the form of its behaviour with temperature.

In an earlier paper (Crawford 1956), experiments were described from which were obtained the real parts of the dynamic shear and bulk moduli of two organic glasses at a frequency of 4 Mc/s and over the temperature range 0 to 100°c, which covered the transition from a solid of viscosity 10^{12} poises to a liquid of viscosity less than 1 poise. It was shown that there was a relaxation in the real part of the bulk modulus which appeared to be at a somewhat lower temperature than the relaxation in the real part of the shear modulus. Poisson's ratio showed an initial reduction with increasing temperature with a fairly sharp minimum before increasing to the maximum value of 0·5.

Further experiments have been carried out on one of the glasses, 2′-hydroxy-2 : 4 : 4 : 6 : 5′-pentamethylflavan (Baker, Curtis and McOmie 1952), at a frequency of 20 Mc/s. As in the earlier experiments, the velocity of longitudinal waves was measured by the optical diffraction method devised by Debye and Sears (1932) and by Lucas and Biquard (1932), but there was a much wider temperature range over which no diffraction pattern was visible. As before also, the shear measurements were made using the pulse technique developed by Mason, Baker, McSkimin and Heiss (1949). The method involves measuring the change in amplitude and phase of a pulse reflected in a quartz bar of rectangular cross section from one of its surfaces when the test material adheres to that surface.

In the earlier work, the amplitude ratio was measured by photographing the display on an oscilloscope of the signal from the quartz crystal receiving the reflected pulse and assuming the oscilloscope system to be linear. The method is open to the objection that there is a considerable time interval between the amplitude measurements with and without the glass adhering to the reflecting surface and therefore the accuracy of the results will be affected by any instabilities in the system. In addition, the results at low temperatures, where the amplitude ratio is of the order of 0·1, are likely to be inaccurate since the received signal with the glass adhering to the surface is very small and the amplifier is approaching saturation when the signal without the glass is received. Both of these objections have been overcome by employing as reference the second quartz bar which was similarly used in the phase change determinations. With no glass adhering to either bar, the ratio between the amplitudes of the first reflected pulses in each bar was measured with an attenuator over the required temperature range. The change in this value with temperature arose from the variation with temperature of the elastic properties of the resin between the quartz crystals and the bars, but the results were reproducible. The ratio was then similarly measured with the organic glass adhering to the reflecting surface of one of the bars and thus the ratio between the amplitudes of the first reflected pulse in one bar with and without the glass adhering to the reflecting surface could be determined. The results were reproducible within approximately ± 0·25 dB.

---

† Now at Imperial Chemical Industries Ltd., Dyestuffs Division, Manchester.

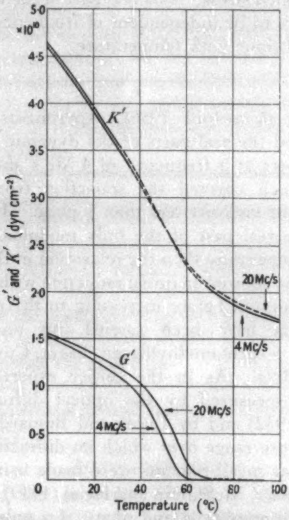


Figure 1. Real parts of shear (G') and bulk moduli (K') plotted against temperature at 4 Mc/s and 20 Mc/s in 2′-hydroxy-2 : 4 : 4 : 6 : 5′-pentamethylflavan.

The measurements at 4 Mc/s were repeated and the real part of the shear modulus was found to be more temperature dependent in the solid state than was indicated by the earlier measurements, falling by almost 16% between 0° and 20°c as compared with 6%. The onset of relaxation is also more gradual. The bulk modulus was calculated from the measurements of the velocity of longitudinal waves and the shear modulus. Figure 1 shows, plotted against temperature, the real parts of the shear and bulk moduli calculated from the new results at 4 Mc/s and 20 Mc/s. The broken line corresponds to the region where no measurements of the velocity of longitudinal waves could be made because there was no visible diffraction pattern.

Measurements of the relaxation time under shear stress at different temperatures (Crawford 1953) and the dynamic shear loss over a wide frequency range, lower than the frequencies used in the present work (Benbow 1954), showed that this glass behaved as if there was one predominant relaxation mechanism with a single relaxation time though there was a slight departure from this behaviour at frequencies higher than that corresponding to the main relaxation time. The present work has confirmed this, in that the main relaxation of the real part of the shear modulus occurs at both frequencies over a comparatively narrow temperature range of about 15°c. Relaxation time measurements (Crawford 1953) indicate that the relaxation times will be of the same order as the periods of the applied stresses at the temperatures at which the value of the real part of the shear modulus begins to fall most rapidly with temperature, namely 45°c at 4 Mc/s and 50°c at 20 Mc/s.

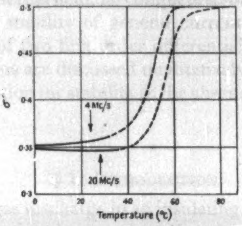


Figure 2. Real part of Poisson's ratio (σ') plotted against temperature at 4 Mc/s and 20 Mc/s in 2′-hydroxy-2 : 4 : 4 : 6 : 5′-pentamethylflavan.

The difference between the values of the real part of the bulk modulus calculated at 20 Mc/s and 4 Mc/s is within the limits of experimental error so that it would seem that the real part of the bulk modulus is independent of frequency. As was reported in the earlier paper (Crawford 1956), there is a transition in the form of its behaviour with temperature, but this transition does not obviously occur at a lower temperature than the transitions in the shear modulus as was apparent in the earlier results. The lack of measurements of the longitudinal velocity over a considerable temperature range makes comparison of the transition regions difficult, but when the bulk modulus results are scaled down to fit the shear modulus results at 0°c, the transitions do appear to occur over approximately the same temperature range and not over a wider range covering the lowest temperatures reached, as seems possible from the curves shown here. This will not be true however at other frequencies if, in fact, the bulk modulus is independent of frequency and the shear transition temperature continues to be dependent on frequency. The transition will thus have to be explained in terms of some volume change which does not depend on the normal viscous properties of the material.

Figure 2 shows the real part of Poisson's ratio plotted against temperature for the two frequencies. The new results at 4 Mc/s do not show the initial reduction in the ratio with temperature as was reported in the earlier paper

(Crawford 1956) but it occurs to some extent at 20 Mc/s. The complete picture of this behaviour will only be seen when a wider frequency range has been covered and the absorption of the longitudinal waves has been measured at the same frequencies.

ACKNOWLEDGMENT

The author would like to express her thanks to Dr. J. H. Andreae for his advice on the electronics involved in this work.

REFERENCES

BAKER, W., CURTIS, R. F., and MCOMIE, J. F. W., 1952, *J. Chem. Soc.*, 1774.

BENBOW, J. J., 1954, *Proc. Phys. Soc.* B, **67**, 120.

CRAWFORD, S. M., 1953, *Proc. Phys. Soc.* B, **66**, 954; 1956, *Ibid.*, **69**, 1312.

DEBYE, P., and SEARS, F. W., 1932, *Proc. Nat. Acad. Sci., Wash.*, **18**, 409.

LUCAS, R., and BIQUARD, P., 1932, *C.R. Acad. Sci., Paris.*, **194**, 2132.

MASON, W. P., BAKER, W. O., MCSKIMIN, H. J., and HEISS, J. H., 1949, *Phys. Rev.*, **75**, 936.

The Influence of an Axial Magnetic Field on the Stability of a Constricted Gas Discharge

By R. J. TAYLER
Atomic Energy Research Establishment, Harwell, Didcot, Berks.

Communicated by B. H. Flowers; *MS. received 25th June* 1957,
and in revised form 23rd August 1957

Abstract. It is shown that a suitable arrangement of axial magnetic fields and conducting walls will stabilize a constricted gas discharge if all the discharge current flows on its surface. For stability the axial field in the discharge exceeds that in the low density region surrounding it and the wall radius cannot exceed the discharge radius by more than a factor of five. In the other extreme case, when the axial current is uniformly distributed across the discharge and the internal field cannot exceed the external field, no complete stability results.

The problem of the stability of general current and field configurations is reduced to the solution of two first order differential equations and a dispersion relation. These equations are discussed qualitatively and we obtain a generalization of a sufficiency condition for stability in the absence of an axial field, which was given in a previous paper.

---

§ 1. Introduction

A constricted gas discharge in an insulating tube is observed to be highly unstable; it wriggles about and reaches the tube walls. Recently published photographs of this behaviour include those of Carruthers and Davenport (1957). Some account of the theory of these instabilities has been given by Kruskal and Schwarzschild (1954) and Tayler (1957 a, to be referred to as I). If a high current constricted discharge is to be obtained in a quasi-stationary state as discussed by Pease (1957), it is necessary to remove these instabilities. The use of external magnetic fields of two types is immediately suggested: (*a*) an axial magnetic field which may, and in general will, be present in both the discharge and the low-density region surrounding it, and (*b*) fields due to eddy currents in highly conducting tube walls. We shall discuss these two possibilities in this paper though they are, of course, not exhaustive. First considerations show that image currents act best on long wavelength instabilities and axial fields on those of short wavelength; can a suitable combination stabilize all wavelengths?

In this paper we are mainly interested in the situation in which a uniform axial magnetic field is initially present in a cylindrical tube containing a slightly conducting gas. At this stage an axial electric field is introduced and the discharge is set up. Further development depends on how rapidly power can be introduced into the system. If the rise in conductivity of the plasma is sufficiently rapid, the axial magnetic field will be largely trapped within the plasma as it contracts. Thus we may hope to reach a quasi-stationary situation, when the discharge is constricted, in which both the axial and azimuthal currents are carried in a thin layer near the discharge surface and there is a virtual separation of axial and azimuthal fields. If the conductivity rises less rapidly, the currents and fields will be more uniformly distributed in the equilibrium configuration. In either case subsequent development due to dissipative processes, because of the finite though large conductivity of both plasma and walls, will lead to interpenetration of the fields. We are thus interested in the stability properties of fairly general configurations of both axial and azimuthal fields.

What we in fact show is that it is possible to obtain completely stable configurations of the first type; that is those in which the currents flow on the discharge surface and the axial magnetic field is mainly trapped within the plasma. These results depend critically on two factors. Stability cannot be obtained without the presence of conducting walls and these walls must be relatively close to the discharge; in the most favourable case the cylindrical wall cannot have a radius greater than five times the discharge radius. Secondly if the discharge is to be noticeably constricted a very high degree of trapping, of the axial field within the plasma, is required; the stable wall position rapidly approaches the discharge as the ratio of internal to external axial field is reduced. We have also been able to consider another extreme case in detail. If the axial current is uniformly distributed across the discharge and the axial field is also uniform, no completely stable configurations exist.

Some of the results contained in this paper have been obtained independently in Russia and reported by Artsimovitch in a talk given at Stockholm in September 1956†. Problems similar in some respects to those considered here have been studied by Dungey and Loughhead (1954) and Roberts (1956) amongst others; although their analysis is similar their emphasis is on properties of magnetic fields rather than current channels.

The remainder of this paper is arranged as follows. In the next section the fundamental equations and assumptions are stated. In § 3 we discuss the properties of completely stable configurations in which all currents are carried on the discharge surface. In § 4 we formulate the equations for general current distributions, and in § 5 obtain a full solution for one special case. In the final section we discuss our results and further problems.

§ 2. ASSUMPTIONS AND BASIC EQUATIONS

In what follows we shall make the following assumptions:

(i) After the initial contraction of the discharge a quasi-equilibrium state is reached. Thus we ignore any instabilities which may occur in this collapse stage.

(ii) The conductivity of the discharge in the quasi-equilibrium state is high enough to be regarded as infinite in stability calculations. This means that the time scale of instabilities which we studied in I is much shorter than the time scale of field penetration due to dissipation. The validity of this assumption can be checked by comparing the times r_0/c_s and $4\pi\sigma r_0^2/c^2$ where r_0 is the discharge radius, c_s the velocity of sound, σ the discharge conductivity and c the velocity of light.

(iii) The wall conductivity may be treated as infinite and we ignore the possible presence of gaps in the conducting walls.

(iv) We consider only equilibrium configurations in infinite cylindrical geometry and with axial symmetry.

---

† I am informed that similar results have been obtained in the United States by M. Kruskal and J. L. Tuck and by M. Rosenbluth and are to be published.

The method of normal mode analysis used in this paper has already been described in I. Perturbations about the equilibrium state are such that any variable takes the form

$$q = q_e + q_1 \exp\{i(m\theta + kz) + \omega t\} \qquad \ldots\ldots(2.1)$$

where k is a real wave number and m is an integer positive or negative. In I all equations were even in m and k so that they could both be taken as positive; in the present work it is found that m, k can both be taken as positive provided that we consider positive and negative values of the axial field. The equilibrium configurations of I are altered by the introduction of axial magnetic fields, and perfectly conducting walls at $r = R_0$. As before displacement currents are neglected and we again state without proof that overstability cannot occur.

Plasma equations.

The plasma equations, restated for reference purposes, are as follows:

$$\rho \frac{d\mathbf{v}}{dt} = -\operatorname{grad} p + \frac{\mathbf{j} \times \mathbf{B}}{c} \qquad \ldots\ldots(2.2)$$

$$\frac{\partial \rho}{\partial t} = -\operatorname{div} \rho \mathbf{v} \qquad \ldots\ldots(2.3)$$

$$\operatorname{curl} \mathbf{B} = \frac{4\pi \mathbf{j}}{c} \qquad \ldots\ldots(2.4)$$

$$\operatorname{div} \mathbf{B} = 0 \qquad \ldots\ldots(2.5)$$

$$\operatorname{curl} \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \qquad \ldots\ldots(2.6)$$

$$\operatorname{div} \mathbf{E} = 4\pi\epsilon \qquad \ldots\ldots(2.7)$$

$$\mathbf{E} + \frac{\mathbf{v} \times \mathbf{B}}{c} = 0 \qquad \ldots\ldots(2.8)$$

and either

$$\frac{1}{p}\frac{dp}{dt} = \frac{\gamma}{\rho}\frac{d\rho}{dt} \qquad \ldots\ldots(2.9)$$

or

$$\operatorname{div} \mathbf{v} = 0. \qquad \ldots\ldots(2.10)$$

Vacuum equations.

In the vacuum the perturbed magnetic field $\mathbf{B}_1$ is both irrotational and solenoidal, leading to

$$B_{1r}{}^{v} = kC_1 K_m{}'(kr) + kC_2 I_m{}'(kr) \qquad \ldots\ldots(2.11)$$

$$B_{1\theta}{}^{v} = \frac{im}{r} C_1 K_m(kr) + \frac{im}{r} C_2 I_m(kr) \qquad \ldots\ldots(2.12)$$

$$B_{1z}{}^{v} = ik\, C_1 K_m(kr) + ik\, C_2 I_m(kr) \qquad \ldots\ldots(2.13)$$

where C_1 and C_2 are constants, arbitrary at present.

Boundary conditions.

The boundary conditions have been stated fully in I. Here it is sufficient to state that for the purpose of obtaining the dispersion relation we need to apply the following two conditions on the perturbed plasma surface:

$$P = p + \mathbf{B}^2/8\pi \text{ continuous,} \qquad \ldots\ldots(2.14)$$

$$B_n \text{ continuous.} \qquad \ldots\ldots(2.15)$$

Only the second condition is required on the rigid conducting wall where it gives

$$\frac{C_1}{C_2} = -\frac{I_m'(kR_0)}{K_m'(kR_0)}. \qquad \ldots\ldots (2.16)$$

The equation of the perturbed plasma surface is

$$r = r_0 + r_1 \exp\{i(m\theta + kz) + \omega t\}$$

and the unit normal vector to this surface is

$$\mathbf{n} = (-1, 0, 0) + \left(0, \frac{imr_1}{r_0}, ikr_1\right) \exp\{i(m\theta + kz) + \omega t\} \ldots\ldots (2.17)$$

§ 3. STABLE CONFIGURATIONS WITH ONLY SURFACE CURRENTS

In equilibrium we have an ideally conducting cylindrical plasma of radius r_0, surrounded by a vacuum, which is itself enclosed by an ideally conducting cylindrical wall at radius R_0. The plasma has uniform density ρ_0, pressure p_0 and ratio of specific heats γ. It carries a magnetic field $(0, 0, B_{\theta 0} b_i)$ and on its surface there is a sheet current $(0, j_{\theta 0}^*, j_{z0}^*)$. In the vacuum there is a magnetic field $(0, B_{\theta 0} r_0/r, B_{\theta 0} b_e)$. The axial fields inside and outside the plasma are therefore expressed as fractions b_i and b_e respectively of the azimuthal field $B_{\theta 0}$ at the plasma surface.

The configuration is in equilibrium provided that

$$\left.\begin{aligned} j_{\theta 0}^* &= \frac{cB_{\theta 0}}{4\pi}[b_i - b_e] \\ j_{z0}^* &= \frac{cB_{\theta 0}}{4\pi} \\ \text{and} \qquad p_0 &= \frac{B_{\theta 0}^2}{8\pi}[1 + b_e^2 - b_i^2]. \end{aligned}\right\} \qquad \ldots\ldots (3.1)$$

We consider perturbations of this equilibrium configuration of the type given by equation (2.1). Equations (2.6) and (2.8) combine to give

$$\mathbf{B}_1 = \frac{ikB_{\theta 0} b_i}{\omega}\left[v_{1r}, v_{1\theta}, v_{1z} + \frac{i}{k}\operatorname{div}\mathbf{v}_1\right]. \qquad \ldots\ldots (3.2)$$

Equations (2.3) and (2.9) give

$$p_1 = -\frac{\rho_0 c_s^2}{\omega}\operatorname{div}\mathbf{v}_1, \qquad \ldots\ldots (3.3)$$

where $c_s^2 = \gamma p_0/\rho_0$, so that c_s is the velocity of sound in the undisturbed plasma.

The third component of equation (2.2) gives

$$v_{1z} = -\frac{ikp_1}{\rho_0 \omega} = \frac{ikc_s^2}{\omega^2}\operatorname{div}\mathbf{v}_1, \qquad \ldots\ldots (3.4)$$

so that equation (3.2) becomes

$$\mathbf{B}_1 = \frac{ikB_{\theta 0} b_i}{\omega}\left[v_{1r}, v_{1\theta}, v_{1z}\left(1 + \frac{\omega^2}{k^2 c_s^2}\right)\right]. \qquad \ldots\ldots (3.5)$$

Then using equations (3.3) and (3.5) in conjunction with the first two components of equation (2.2) we obtain expressions for v_{1r} and $v_{1\theta}$ in terms of v_{1z} :

$$v_{1r} = -\frac{iDv_{1z}[k^2 + \omega^2/c_H^2 + \omega^2/c_s^2]}{k[k^2 + \omega^2/c_H^2]}, \qquad \ldots\ldots (3.6)$$

$$v_{1\theta} = \frac{mv_{1z}[k^2 + \omega^2/c_H^2 + \omega^2/c_s^2]}{kr[k^2 + \omega^2/c_H^2]}, \qquad \ldots\ldots (3.7)$$

where D stands for d/dr and $c_H{}^2 = B_{\theta 0}{}^2 b_i{}^2/4\pi\rho_0$, so that c_H is the hydromagnetic velocity in the undisturbed plasma. Using equations (3.4), (3.6) and (3.7) we now obtain one equation for v_{1z}:

$$rD[rDv_{1z}] = v_{1z}\left[m^2 + \frac{(k^2+\omega^2/c_H{}^2)(k^2+\omega^2/c_s{}^2)}{[k^2+\omega^2/c_H{}^2+\omega^2/c_s{}^2]}\,r^2\right]. \qquad \ldots\ldots(3.8)$$

This is a modified Bessel equation and its solution is

$$\left.\begin{aligned} v_{1z} &= CI_m(\alpha r)\\ v_{1\theta} &= \frac{m[k^2+\omega^2/c_H{}^2+\omega^2/c_s{}^2]}{k\,[k^2+\omega^2/c_H{}^2]}\,C\,\frac{I_m(\alpha r)}{r}\\ v_{1r} &= \frac{-i\alpha[k^2+\omega^2/c_H{}^2+\omega^2/c_s{}^2]}{k\,[k^2+\omega^2/c_H{}^2]}\,C\,I_m{}'(\alpha r) \end{aligned}\right\} \qquad \ldots\ldots(3.9)$$

where the expressions for $v_{1\theta}$ and v_{1r} have been obtained from equations (3.7) and (3.6), C is an arbitrary constant and

$$\alpha^2 = \frac{(k^2+\omega^2/c_H{}^2)(k^2+\omega^2/c_s{}^2)}{[k^2+\omega^2/c_H{}^2+\omega^2/c_s{}^2]}.$$

The boundary conditions can now be applied on the perturbed plasma boundary, using the vacuum solutions of § 2, and we obtain

$$\alpha k r_0{}^2\left[b_i{}^2 + \frac{\gamma(\omega^2/c_s{}^2)(1+b_e{}^2-b_i{}^2)}{2(k^2+\omega^2/c_s{}^2)}\right]\frac{I_m(\alpha r_0)}{I_m{}'(\alpha r_0)}$$
$$= kr_0 + (m+b_e kr_0)^2\left[\frac{K_m(kr_0)\,I_m{}'(kR_0) - I_m(kr_0)\,K_m{}'(kR_0)}{K_m{}'(kr_0)I_m{}'(kR_0) - I_m{}'(kr_0)K_m{}'(kR_0)}\right] \qquad \ldots\ldots(3.10)$$

It is convenient to express this equation in terms of non-dimensional variables. Thus we make the substitutions

$$kr_0 = X_0,\quad \alpha r_0 = U_0,\quad \omega r_0/c_s = W_0 \quad\text{and}\quad R_0/r_0 = \Lambda. \qquad \ldots\ldots(3.11)$$

Then equation (3.10) becomes

$$X_0\,U_0\left[b_i{}^2 + \frac{\gamma W_0{}^2(1+b_e{}^2-b_i{}^2)}{2(X_0{}^2+W_0{}^2)}\right]\frac{I_m(U_0)}{I_m{}'(U_0)}$$
$$= X_0 + (m+b_e X_0)^2\left[\frac{K_m(X_0)\,I_m{}'(\Lambda X_0) - I_m(X_0)\,K_m{}'(\Lambda X_0)}{K_m{}'(X_0)\,I_m{}'(\Lambda X_0) - I_m{}'(X_0)\,K_m{}'(\Lambda X_0)}\right]. \qquad \ldots\ldots(3.12)$$

In stable configurations, it must be possible to choose values of the three parameters b_i, b_e and Λ so that for no positive values of X_0 and for no positive integral values of m can equation (3.12) have a positive root for $W_0{}^2$. As mentioned in § 2, in order to allow for all possible perturbations we must consider positive and negative values of b_i and b_e. In fact equation (3.12) is even in b_i, and we shall find that we need only consider negative b_e as positive b_e always gives greater stability. b_i, b_e and Λ are not completely free parameters as they must satisfy the inequalities

$$1 + b_e{}^2 \geqslant b_i{}^2, \quad \Lambda \geqslant 1. \qquad \ldots\ldots(3.13)$$

The first inequality follows from the third of equations (3.1) since the plasma pressure must be positive.

The right-hand side of equation (3.12) is independent of W_0; the left-hand side varies between $b_i{}^2 X_0{}^2 I_m(X_0)/I_m{}'(X_0)$ and infinity as $W_0{}^2$ varies between zero and infinity and it is never less than $b_i{}^2 X_0{}^2 I_m(X_0)/I_m{}'(X_0)$. Thus equation (3.12) cannot have a solution for positive W_0^2 if

$$X_0+(m+b_e X_0)^2\left[\frac{K_m(X_0)\,I_m'(\Lambda X_0)-I_m(X_0)\,K_m'(\Lambda X_0)}{K_m'(X_0)\,I_m'(\Lambda X_0)-I_m'(X_0)\,K_m'(\Lambda X_0)}\right]<\frac{b_i^2\,X_0^2 I_m(X_0)}{I_m'(X_0)}. \qquad \ldots\ldots(3.14)$$

This is both a necessary and sufficient condition for stability, so that if inequality (3.14) fails to be satisfied for any m, X_0 the configuration is unstable. The inequality has two obvious properties. There cannot be complete stability in the absence of a trapped axial field ($b_i=0$); for then the right-hand side of inequality (3.14) is zero while the left-hand side is certainly positive for negative b_e and m and X_0 satisfying $m+b_e X_0=0$. Secondly if the inequality is satisfied for any negative b_e it is also satisfied for the corresponding positive b_e; the additional term

$$4mb_e X_0\left[\frac{K_m(X_0)\,I_m'(\Lambda X_0)-I_m(X_0)\,K_m'(\Lambda X_0)}{K_m'(X_0)\,I_m'(\Lambda X_0)-I_m'(X_0)\,K_m'(\Lambda X_0)}\right]$$

is negative. Thus as mentioned above we need only consider negative b_e. The instability, in the absence of a trapped field, with the wave number satisfying $m+b_e X_0=0$, occurs when the perturbation helix and the equilibrium magnetic field helix at the plasma surface coincide.

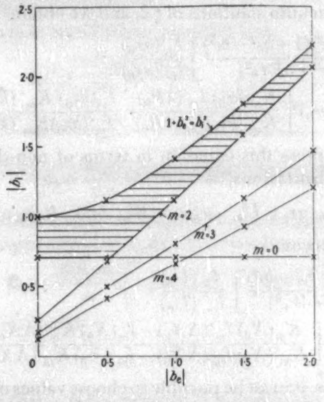


Figure 1. Stability diagram for surface current, no conducting walls. Stability is obtained for values of $|b_e|$, $|b_i|$ lying above the curve corresponding to a given value of m and below the curve $1+b_e^2=b_i^2$. Complete stability for $m\neq 1$ is obtained in the shaded region.

We consider the behaviour of the inequality (3.14) first taking into account axial magnetic fields but no conducting walls ($\Lambda=\infty$). In this case we find it is possible to stabilize both the $m=0$ modes and those with $m>1$ for suitable values of b_i and b_e, but that without conducting walls it is not possible to stabilize the $m=1$ modes. For given m and $|b_e|$ there is a critical value of $|b_i|$ above which complete stability is obtained; at the same time $|b_i|$ must satisfy inequality (3.13). The results are shown in table 1 and also graphically in figure 1. In the absence of conducting walls the critical value of $|b_i|$ for $m=0$ is independent of $|b_e|$. It can be seen that the curve $1+b_e^2=b_i^2$ and the critical points for $m=0$ and $m=2$

define a region (shaded in figure 1) in which all but the $m=1$ modes are stable. In this region $m=1$ disturbances of high wave number will be stable. In table 2 we show the critical wave numbers above which stability occurs for a series of points on the upper and lower boundaries of the shaded region in figure 1.

Table 1. Surface Current. Critical Values of $|b_i|$ for Stability

$m \backslash |b_e|$	0·0	0·5	1·0	1·5	2·0
0	0·707	0·707	0·707	0·707	0·707
2	0·259	0·676	1·122	1·592	2·073
3	0·169	0·492	0·801	1·130	1·468
4	0·126	0·409	0·660	0·926	1·202

Table 2. Surface Current. Critical Wave Numbers for Stability, $m=1$

$|b_e|$	0·0	0·5	1·0	1·5	2·0
$|b_i|$	1·000	1·118	1·414	1·803	2·236
X_0	0·450	0·869	0·736	0·571	0·455
$|b_i|$	0·707	0·707	1·122	1·592	2·073
X_0	1·025	1·886	0·991	0·661	0·497

If conducting walls are now introduced we find that the $m=1$ modes can also be completely stabilized, though only if the walls are quite close to the discharge. The values of $|b_i|$, $|b_e|$ for which complete stability can be obtained and the corresponding values of Λ are shown in tables 3 and 4 and figure 2. Table 3 shows the values of Λ below which complete stability for $m=1$ can be obtained if $|b_i|$ has its maximum value $(1+b_e^2=b_i^2)$. It can be seen that for $b_e=0$, Λ has a maximum value of only 5, and that as $|b_e|$ increases from zero the value of Λ rapidly decreases. For given $|b_e|$ as $|b_i|$ decreases the critical value of Λ decreases. Λ must be greater than 1 and the minimum values of $|b_i|$ for which stability is obtained are shown in table 4. For other values of Λ between 1 and 5 we can obtain similar stability curves in the $(|b_i|, |b_e|)$ plane and several of these are shown in figure 2.

Table 3. Surface Current. Critical Wall Radii for Stability, $m=1$

$|b_e|$	0·0	0·1	0·2	0·5	1·0
$|b_i|$	1·000	1·005	1·020	1·118	1·414
Λ	5·00	3·31	2·62	1·81	1·39

Table 4. Surface Current. Minimum Values of $|b_i|$ for Stability, $m=1$, $\Lambda=1$

$|b_e|$	0·0	0·2	0·5	1·0	1·5	2·0
$|b_i|$	0·000	0·429	0·683	1·114	1·580	2·061

The region of complete stability is mainly determined by the $m=1$ behaviour but the $m=0$ disturbances must also be considered. With finite Λ stability with $m=0$ is obtained for $|b_i|$ greater than the value given by

$$b_i^2 = \tfrac{1}{2} - b_e^2/(\Lambda^2-1). \qquad \ldots\ldots(3.15)$$

Solutions of equation (3.15) for several values of Λ, $|b_e|$ are given in table 5 and plotted in figure 2.

Table 5. Surface Current. Minimum Values of $|b_1|$ for Stability, $m=0$

$\Lambda \backslash |b_e|$	0·0	0·5	1·0	1·5	2·0
5·0	0·707	0·700	0·677	0·637	0·577
4·0	0·707	0·695	0·658	0·592	0·483
3·0	0·707	0·685	0·612	0·468	0·000
2·0	0·707	0·645	0·408	0·000	0·000
1·5	0·707	0·548	0·000	0·000	0·000

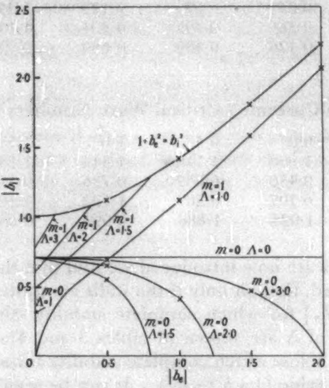


Figure 2. Stability diagram for surface current, with conducting walls. Stability occurs in the region above the curve marked with the given values of m and Λ and below the curve $1+b_e^2=b_1^2$. Complete stability for given Λ occurs when a point lies in the stability region for both $m=0$ and $m=1$.

Combining the results for $m=0$ and $m=1$, complete stability regions for four values of Λ (1·0, 1·5, 2·0, 3·0) are shown in figure 2. Thus as stated in § 1 complete stability can be obtained, but with two critical provisions: (*a*) wall radius is not too greatly different from discharge radius; (*b*) $|b_e|$ is very small unless Λ is very close to 1.

There is one other condition we may wish to apply. If the constricted discharge is to be a major element in the configuration the heat energy stored in the discharge should not be negligible compared with the total magnetic energy. Clearly it will be negligible if we go to a configuration in which Λ is approximately equal to unity and $|b_e|(\simeq|b_1|)\gg 1$; in this case the axial current and constriction are of relatively minor importance. In table 6 we tabulate the plasma energy ϵ_p as a fraction of the total energy for sets of ($|b_1|$, $|b_e|$, Λ) for which stability is obtained.

Table 6. Stable Configurations. Fraction of Energy Contained in Plasma

Λ	1·0	1·0	1·0	1·5	1·5	2·0	2·0	3·0	3·0	5·0
$|b_e|$	0·0	1·0	2·0	0·0	0·5	0·0	0·2	0·0	0·1	0·0
$|b_1|$	0·0	1·114	2·061	0·707	0·959	0·707	0·840	0·777	0·947	1·0
ϵ_p	1·000	0·478	0·210	0·364	0·195	0·285	0·185	0·175	0·051	0·000

§ 4. CONFIGURATIONS WITH CURRENTS NOT CONFINED TO THE SURFACE

We consider now the case which arises if the initial rate of current rise of the discharge is not high enough for the surface currents and separated fields of the last section to be realized; similarly these conditions will be obtained when the initially separated fields have had time to interpenetrate. We obtain the equations for the case of a completely general distribution of axial currents and magnetic fields, but in this paper we consider only the case of an incompressible fluid. This restriction is made purely for algebraic simplicity since in the problems studied in I stability criteria were insensitive to this assumption. There is no difference in principle and the equations for the compressible fluid can be treated by the methods of this section.

We assume that in equilibrium we have an incompressible ideally conducting fluid of density ρ_0, forming a cylinder of radius r_0. The fluid carries a magnetic field $\mathbf{B}_0$ and current $\mathbf{j}_0$ and has pressure p_0, where $\mathbf{B}_0$, $\mathbf{j}_0$ and p_0 have the radial dependence given by

$$\mathbf{B}_0 = B_{\theta 0}\,[0, f(r)/f(r_0),\ b\,g(r)/g(r_0)], \qquad \ldots\ldots(4.1)$$

$$\mathbf{j}_0 = \frac{cB_{\theta 0}}{4\pi}\,[0, bg'/g_0, (rf)'/rf_0] \qquad \ldots\ldots(4.2)$$

and

$$p_0 = \frac{B_{\theta 0}^2}{8\pi}\left[b^2(1-g^2/g_0^2) + 1 - f^2/f_0^2 + 2\int_r^{r_0} (f^2/rf_0^2)\,dr\right] \qquad \ldots\ldots(4.3)$$

where the prime denotes differentiation with respect to r. In equations (4.2) and (4.3) and in what follows we write f for $f(r)$ and f_0 for $f(r_0)$ and similarly for the other functions. We have chosen the constant in equation (4.3) so that the pressure vanishes at the plasma surface, and the plasma is surrounded by a vacuum carrying a magnetic field $(0, B_{\theta 0} r_0/r, B_{\theta 0} b)$; as mentioned in I it is more satisfactory to assume that the outer region has a small but finite density, but the stability criteria are only slightly affected by such an assumption.

A perturbation of the form (2.1) is now applied, and from equations (2.6), (2.8) and (2.10) we find that the perturbed magnetic field can be expressed in terms of the perturbed velocities in the form

$$\begin{aligned}\mathbf{B}_1 = \frac{iB_{\theta 0}}{\omega r}\Bigg[&\left(\frac{mf}{f_0} + \frac{bkrg}{g_0}\right) v_{1r}, \left(\frac{mf}{f_0} + \frac{bkrg}{g_0}\right) v_{1\theta} + \frac{ir^2}{f_0}\left(\frac{f}{r}\right)' v_{1r},\\ &\left(\frac{mf}{f_0} + \frac{bkrg}{g_0}\right) v_{1z} + \frac{ibrg'}{g_0}\, v_{1r}\Bigg]. \qquad \ldots\ldots(4.4)\end{aligned}$$

The second and third components of equation (2.2) (using equations (4.1), (4.2), (4.4) and (2.4)) now give two equations for p_1 in terms of the components of v_1:

$$\begin{aligned}&v_{1\theta}\left[Y_0^2 + \frac{bkr_0^2 g}{rg_0}\left(\frac{mf}{f_0} + \frac{bkrg}{g_0}\right)\right] - v_{1z}\left[\frac{bmr_0^2 g}{r^2 g_0}\left(\frac{mf}{f_0} + \frac{bkrg}{g_0}\right)\right]\\ &\quad + \frac{imY_0^2}{\rho_0 \omega r}\, p_1 - iv_{1r}\left[\frac{r_0^2 (rf)'}{r^2 f_0}\left(\frac{mf}{f_0} + \frac{bkrg}{g_0}\right) + \frac{b^2 m r_0^2 gg'}{rg_0^2} - \frac{bkr_0^2 rg}{f_0 g_0}\left(\frac{f}{r}\right)'\right] = 0, \qquad \ldots\ldots(4.5)\end{aligned}$$

and

$$\begin{aligned}&v_{1z}\left[Y_0^2 + \frac{mr_0^2 f}{r^2 f_0}\left(\frac{mf}{f_0} + \frac{bkrg}{g_0}\right)\right] - v_{1\theta}\left[\frac{kr_0^2 f}{rf_0}\left(\frac{mf}{f_0} + \frac{bkrg}{g_0}\right)\right]\\ &\quad + \frac{ikY_0^2}{\rho_0 \omega}\, p_1 - iv_{1r}\left[\frac{br_0^2 g'}{rg_0}\left(\frac{mf}{f_0} + \frac{bkrg}{g_0}\right) + \frac{b^2 m r_0^2 fg'}{rf_0 g_0} - \frac{kr_0^2 rf}{f_0^2}\left(\frac{f}{r}\right)'\right] = 0, \qquad \ldots\ldots(4.6)\end{aligned}$$

where we have introduced the dimensionless variable $Y_0{}^2 = 4\pi\rho_0\,\omega^2 r_0{}^2/B_{\theta 0}{}^2$. We eliminate p_1 between equations (4.5) and (4.6) to obtain

$$[kr\,v_{1\theta} - mv_{1z}]\left[Y_0{}^2 + \frac{r_0{}^2}{r^2}\left(\frac{mf}{f_0} + \frac{bkrg}{g_0}\right)^2\right] - 2ikrv_{1r}\frac{r_0{}^2 f}{r^2 f_0}\left(\frac{mf}{f_0} + \frac{bkrg}{g_0}\right) = 0. \qquad \ldots\ldots(4.7)$$

If this is combined with equation (2.10) we obtain an equation between v_{1r} and $v_{1\theta}$

$$m(rv_{1r})' + \frac{2k^2r^2\dfrac{r_0{}^2 f}{r^2 f_0}\left(\dfrac{mf}{f_0} + \dfrac{bkrg}{g_0}\right)}{\left[Y_0{}^2 + \dfrac{r_0{}^2}{r^2}\left(\dfrac{mf}{f_0} + \dfrac{bkrg}{g_0}\right)^2\right]}v_{1r} + iv_{1\theta}[m^2 + k^2r^2] = 0. \qquad \ldots\ldots(4.8)$$

The first component of the equation of motion (2.2) yields a highly complicated equation between p_1, v_{1r}, $v_{1\theta}$, v_{1z} and their first derivatives. However p_1 and v_{1z} can be eliminated using equations (4.5) and (4.6) and we obtain a further equation between v_{1r} and $v_{1\theta}$:

$$\begin{aligned}\frac{2r_0{}^2 f}{r^2 f_0}\left(\frac{mf}{f_0} + \frac{bkrg}{g_0}\right)(rv_{1r})' &+ m\left[Y_0{}^2 + \left(\frac{mf}{f_0} + \frac{bkrg}{g_0}\right)\left(\frac{2r_0{}^2 f'}{mrf_0} - \frac{2r_0{}^2 f}{mr^2 f_0} + \frac{mr_0{}^2 f}{r^2 f_0} + \frac{bkr_0{}^2 g}{rg_0}\right)\right.\\ &\left. + \frac{2bkr_0{}^2 fg'}{mf_0 g_0}\right]v_{1r} + i\left[Y_0{}^2 + \frac{r_0{}^2}{r^2}\left(\frac{mf}{f_0} + \frac{bkrg}{g_0}\right)^2\right](rv_{1\theta})'\\ &+ i\frac{2r_0{}^2}{r}\left(\frac{mf}{f_0} + \frac{bkrg}{g_0}\right)\left(\frac{mf'}{f_0} + \frac{bkrg'}{g_0}\right)v_{1\theta} = 0. \qquad \ldots\ldots(4.9)\end{aligned}$$

We now make the substitutions

$$\left.\begin{aligned} &mv_{1r} = \phi \;\text{ and }\; m\beta\,v_{1r} + iv_{1\theta} = \chi\\ \text{where}\qquad &\beta = \frac{2r_0{}^2 f}{r^2 f_0}\left[\frac{mf}{f_0} + \frac{bkrg}{g_0}\right]\Big/ m\left[Y_0{}^2 + \frac{r_0{}^2}{r^2}\left(\frac{mf}{f_0} + \frac{bkrg}{g_0}\right)^2\right].\end{aligned}\right\} \qquad \ldots\ldots(4.10)$$

The variables ϕ and χ are not immediately relevant to the case $m = 0$ but that case can simply be treated separately. Using (4.10), equations (4.8) and (4.9) become

$$(r\phi)' + (m^2 + k^2r^2)\chi - m^2\beta\phi = 0 \qquad \ldots\ldots(4.11)$$

$$\begin{aligned}(r\chi)' &+ m\beta\left[\frac{rf_0}{f}\left(\frac{mf'}{f_0} + \frac{bkrg'}{g_0}\right)\right]\chi\\ &+ \left[1 - \left\{\frac{m(rf' - f)}{f_0}\Big/\left(\frac{mf}{f_0} + \frac{bkrg}{g_0}\right)\right\}\beta - m^2\beta^2\right]\phi = 0. \qquad \ldots\ldots(4.12)\end{aligned}$$

We can now apply the boundary conditions (2.14), (2.15) at the perturbed plasma surface and obtain the dispersion relation between m, k and ω in terms of the values ϕ_0, χ_0 of ϕ and χ at $r = r_0$. Thus

$$\frac{\chi_0}{\phi_0} = -\frac{(m + bkr_0)^2}{kr_0[Y_0{}^2 + (m + bkr_0)^2]}\,\frac{[K_m(kr_0)\,I_m{}'(kR_0) - I_m(kr_0)\,K_m{}'(kR_0)]}{[K_m{}'(kr_0)\,I_m{}'(kR_0) - I_m{}'(kr_0)\,K_m{}'(kR_0)]}. \qquad \ldots\ldots(4.13)$$

We again introduce dimensionless variables (3.11) to obtain

$$\frac{\chi_0}{\phi_0} = -\frac{(m + bX_0)^2}{X_0[Y_0{}^2 + (m + bX_0)^2]}\,\frac{[K_m(X_0)\,I_m{}'(\Lambda X_0) - I_m(X_0)\,K_m{}'(\Lambda X_0)]}{[K_m{}'(X_0)\,I_m{}'(\Lambda X_0) - I_m{}'(X_0)\,K_m{}'(\Lambda X_0)]}. \qquad \ldots\ldots(4.14)$$

Although the equations (4.11) and (4.12) and the dispersion relation (4.14) are relatively simple, it is clear that any general solution of them must be obtained numerically. However we can make some general comments. We are interested in the possible existence of functions f and g and constants b and Λ so that for no values of X_0 and m can the dispersion relation have solutions for positive values of Y_0^2. Such functions f and g must yield values of p_0 which are consistent with the idea of a constricted gas discharge and give field configurations which might be expected to arise naturally.

If $Y_0^2>0$, the right-hand side of equation (4.14) is positive; thus ϕ_0 and χ_0 must have the same sign. In general ϕ and χ have opposite signs at $r=0$. (The only non-singular exception is when there is a uniform distribution of axial current near the discharge centre.) ϕ cannot change sign before χ and χ cannot change sign while

$$1-\left\{\frac{m(rf'-f)}{f_0}\bigg/\left(\frac{mf}{f_0}+\frac{bkrg}{g_0}\right)\right\}\beta-m^2\beta^2\geqslant 0. \qquad \ldots\ldots(4.15)$$

This condition can be reduced to the form

$$\left(m+\frac{bkrgf_0}{fg_0}\right)^2\geqslant\frac{2(rf'+f)}{f}=\frac{2r}{I_z}\frac{dI_z}{dr} \qquad \ldots\ldots(4.16)$$

where $I_z(r)$ is the total axial current contained within radius r. Equation (4.16) is a generalization of one that was given in I. Thus in the absence of an axial field instability cannot occur if

$$m^2\geqslant\left(\frac{2r}{I_z}\frac{dI_z}{dr}\right)_{\max} \qquad \ldots\ldots(4.17)$$

where the suffix means that we must take the maximum value of the quantity between $r=0$ and $r=r_0$. Note that this sufficiency condition can give no information about modes with $m<2$ as for non-singular current distributions

$$\left(\frac{2r}{I_z}\frac{dI_z}{dr}\right)_{\max}\geqslant 4.$$

It is clear that the inequality cannot be as useful when the axial field is included; for, if b is negative and if f and g are given, there must be combinations of m and k for which it fails. There are however some values of m and k for which there must be stability, and for other values of m and k the inequality may only fail to be satisfied for a narrow range in r; in this case χ must change sign within this range if instability can occur. The inequality itself is not a very restrictive condition and it can be shown that in interesting cases χ is still increasing at the time (4.16) fails to be satisfied; thus it must have a turning value before it can vanish. These general results should guide the numerical work which we hope will follow the present paper.

§ 5. UNIFORM AXIAL FIELD AND CURRENT

Now we turn our attention to one special case for which a full solution can be obtained. If the axial field and current are both uniform ($g\propto 1$, $f\propto r$) equation (4.12) reduces to the simpler form

$$(r\chi)'+m^2\beta\chi+(1-m^2\beta^2)\phi=0 \qquad \ldots\ldots(5.1)$$

where β is now a constant. We generalize the problem slightly to allow for differing internal and external axial fields, b_i, b_e ($b_i^2<b_e^2$). Thus

$$\beta=2(m+b_ikr_0)/m\,[Y_0^2+(m+b_ikr_0)^2]$$

and the dispersion relation becomes

$$\frac{\chi_0}{\phi_0}=-\frac{(m+b_e X_0)^2}{X_0[Y_0{}^2+(m+b_1 X_0)^2]}\,\frac{[K_m(X_0)\,I_m'(\Lambda X_0)-I_m(X_0)\,K_m'(\Lambda X_0)]}{[K_m'(X_0)\,I_m'(\Lambda X_0)-I_m'(X_0)\,K_m'(\Lambda X_0)]}. \qquad \ldots\ldots(5.2)$$

Table 7. Volume Current. Equal Internal and External Fields, no Walls. Table of Wave Numbers and Corresponding Growth Rates.

X_0 b	0·0	0·5	−0·5	1·0	−1·0	2·0	−2·0	5·0	−5·0
0·05								−0·6072	0·3796
0·10							0·3403	−1·3851	0·5062
0·15								−2·2175	0·3779
0·20				−0·3649		−0·9531	0·5059		0·0000
0·25		−0·1450			0·4315				−0·0302
0·30						−1·5089	0·4997		−0·1730
0·35									−0·4290
0·40				−0·7054		−2·0876	0·3269		
0·50	0·2481	−0·2169	0·5088		0·5575		0·0000		
0·60				−1·0455			0·0212		
0·70							−0·0193		
0·75		−0·2826			0·3959				
0·80				−1·4083					
1·00	0·4670	−0·3645	0·6090	−1·8111	0·0000				
1·20					0·0778				
1·40					0·1067				
1·50		−0·6054	0·4095						
1·60					0·0834				
1·80					0·0048				
2·00	0·6935	−0·9650	0·0000						
2·50			0·2053						
3·00			0·3447						
3·50	0·8319		0·4023						

In obtaining these results care is required to find the largest root of the equation (5.2) for $Y_0{}^2$ (Tayler 1957 b).

Except that the definition of β has been altered the equations (4.11) and (5.1) are exactly those found in I in the absence of an axial field and their solution is

$$\phi \propto \frac{m^2\beta J_m[(m^2\beta^2-1)^{1/2}kr]}{(m^2\beta^2-1)\,kr}+\frac{J_m'[(m^2\beta^2-1)^{1/2}\,kr]}{(m^2\beta^2-1)^{1/2}},$$

$$\chi \propto \frac{J_m[(m^2\beta^2-1)^{1/2}\,kr]}{kr}.$$

We have considered full solutions for $m=1$ which was the most difficult mode to stabilize for surface currents. It is immediately fairly easily seen that there can be no complete stability. At the point $1+b_e X_0=0$ the largest root for $Y_0{}^2$ corresponds to $\beta=1$. Thus $Y_0{}^2=1-b_1{}^2/b_e{}^2$, and this is positive unless $b_1=b_e$. If $b_1=b_e$ a perturbation expansion shows that $Y_0{}^2$ is decreasing with X_0 at this point and is positive for smaller X_0. Thus stability cannot be expected for completely penetrated fields. Qualitatively the results are only slightly affected by the presence of conducting walls although some stability is produced at long wavelengths. Results for $m=1$ are shown in tables 7 and 8 and figure 3. In table7 and figure 3 are given results for $\Lambda=\infty$ (no walls) and several values of $b_1(=b_e)$. In table 8 are shown results for $b_e=-1$ and several values of b_1 and Λ. The results may be expected to be qualitatively true for a compressible fluid although the discontinuity in the growth rate against wave number curve at $b_1X_0 = -1$ should be removed.

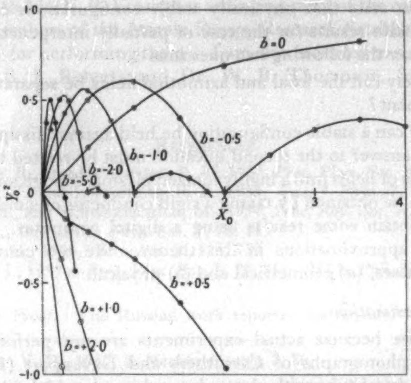


Figure 3. Volume current. Plot of growth rate against wave number for the case of equal internal and external fields. Curves are plotted for eight values of $b_e (= b_1)$; they show the square of the dimensionless growth rate ($Y_0^2 = 4\pi\rho_0\omega^2 r_0^2 / B_{\theta 0}^2$) as a function of the dimensionless wave number ($X_0 = kr_0$).

Table 8. Volume Current. Internal and External Fields and Conducting Walls. Table of Wave Numbers and Corresponding Growth Rates; $b_e = -1$

		X_0										
b_1	Λ	0·0	0·2	0·4	0·6	0·8	1·0	1·2	1·4	1·6	1·8	2·0
−1	∞	0·0000	0·3673	0·5449	0·5240	0·3341	0·0000	0·0778	0·1067	0·0834	0·0048	<0
−1	2	−0·6667	−0·0204	0·3795	0·4741	0·3260	0·0000	0·0775	0·1061	0·0824		<0
0	∞	0·0000	0·4063	0·6970	0·8829	0·9738	1·0000	0·9788	0·9240	0·8485	0·7667	0·6935
0	2	−0·6667	0·0134	0·5334	0·8314	0·9656	1·0000	0·9756	0·9164	0·8387	0·7573	0·6861

		$1/X_0$										
		1·0	0·9	0·8	0·7	0·6	0·5	0·4	0·3	0·2	0·1	0·0
0	∞	1·0000	0·9931	0·9677	0·9140	0·8212	0·6935	0·6054	0·6575	0·8044	0·9443	1·0000
0	2	1·0000	0·9919	0·9634	0·9060	0·8112	0·6861	0·6034	0·6574	0·8044	0·9443	1·0000

§ 6. CONCLUSIONS AND FURTHER PROBLEMS

We can summarize the mathematical results of the last two sections by stating that stable configurations may exist with the following properties:

(*a*) The conducting walls must be relatively close to the plasma and most favourable results are obtained if the axial magnetic field is mainly trapped within the plasma.

(*b*) There is thus no stability against small disturbances for highly constricted discharges of the type considered by Pease (1957); we have not however ruled out the possibility that such instabilities could be limited at finite amplitude.

(*c*) The most stable configurations in the absence of external fields are not the most easily stabilized in their presence; thus a uniform current distribution was the most stable configuration discussed in I but the fields considered in the present paper cannot make it completely stable.

Before we can state that physically stable configurations can exist it is of importance to obtain results for the case of partially interpenetrated fields, and to be able to answer the following two questions :

(*a*) How closely can the axial and azimuthal fields be separated in an experimental arrangement?

(*b*) How long can a stable configuration be held against dissipative processes? In principle any answer to the second question must be related to a study of the rate of penetration of fields into a highly ionized plasma, though an order of magnitude estimate can be obtained by taking a rigid conductor of similar conductivity. It is hoped to obtain some results using a digital computer. There are still obviously many approximations in this theory. We can conveniently divide them into two classes, (*a*) geometrical and (*b*) physical.

Geometrical approximations.

(*a*) These arise because actual experiments are not performed in infinite cylinders. The photographs of Carruthers and Davenport (1957) show discharges in a cylindrical tube with electrodes and in a toroidal tube. In addition metal walls may not be continuous but may contain slits for the introduction of fields.

No direct attempt is made in this paper to discuss the possibility of new instabilities introduced by toroidal geometry and the curvature of the equilibrium current channel. An attempt can be made to approach finite geometry by applying periodic boundary conditions to results already obtained in an infinite cylinder; this places an upper limit on the possible instability wavelengths. Thus we can formally obtain a completely stable configuration for the case of surface currents, provided that the critical wavelength below which there is stability (shown in table 2) is greater than the length of the system L. For X_0 to be small enough for the resultant dimension to be practicable we must be in the part of the stability diagram (figure 1) where $|b_e| \gg 1$, $|b_i| \simeq |b_e|$. Under these conditions there is complete stability provided that $b_i > L/2\pi r_0$.

This is however a situation in which the external field considerably exceeds the discharge field and not one in which we are strictly interested in this paper.

The problem of periodic slits in an infinite conducting wall can be treated by the method of this paper but the resulting dispersion relation is in the form of a very complicated infinite determinant. However, qualitatively it appears that narrow slits should not have a serious effect on an otherwise completely stabilized discharge. Even in the absence of walls only long wavelength instabilities remain; the narrow slits alter these normal modes by the introduction of higher harmonics of wavelength comparable with the slit width and these are prevented from growing by the trapped axial field.

Physical approximations.

(*b*) Obvious physical approximations are the neglect of all dissipative terms and especially the artificial division of the fluid into regions of infinite and zero electrical conductivity. One such neglected effect which might lead to slightly enhanced stability is the anisotropic electrical conductivity of the discharge; this causes an axial magnetic field to increase towards the discharge centre.

ACKNOWLEDGMENTS

I am grateful to Mr. E. B. Fossey for arranging most of the calculations and to him and other members of the Atomic Energy Research Establishment, Harwell, computing group for performing them. I am also indebted to several colleagues, especially Miss S. J. Roberts and Dr. W. B. Thompson, for many helpful discussions.

REFERENCES

CARRUTHERS, R., and DAVENPORT, P. A., 1957, *Proc. Phys. Soc.* B, **70**, 49.
DUNGEY, J. W., and LOUGHHEAD, R. E., 1954, *Aust. J. Phys.*, **7**, 5.
KRUSKAL, M., and SCHWARZSCHILD, M., 1954, *Proc. Roy. Soc.* A, **223**, 348.
PEASE, R. S., 1957, *Proc. Phys. Soc.* B, **70**, 11.
ROBERTS, P. H., 1956, *Astrophys. J.*, **124**, 430.
TAYLER, R. J., 1957 a, *Proc. Phys. Soc.* B, **70**, 31; 1957 b, *Phil. Mag.*, **2**, 33.

Note Added in Proof. The Russian work reported by Artsimovitch has now been published in English:

SHRAFRANOV, V. D., 1957, *J. Nuclear Energy*, **2**, 86.

The Polarization Reversal Process in Guanidine Aluminium Sulphate Hexahydrate

By M. PRUTTON

Birkbeck College Computational Laboratory, University of London

MS. received 20th June 1957

Abstract. The activation fields of guanidine aluminium sulphate hexahydrate found by three methods are compared. Only over part of the range of fields used is the polarization reversal process characterized by a simple exponential dependence on the field as in $BaTiO_3$. The coercivity and rate of change of polarization are shown to be dependent more upon the rate of change of switching field than upon the frequency or repetition rate.

---

§ 1. Introduction

The switching properties of the ferroelectric crystals of guanidine aluminium sulphate hexahydrate have been briefly reported by Holden, Merz, Remeika and Matthias (1956) but no account of the dynamics of domain growth and nucleation in this material or any of its class has yet been published. The study of this reversal process has been carried out here by three methods : the investigation of the current flowing in the crystal due to alternate positive and negative square pulses, the variation of the coercivity with the frequency of a sinusoidally varying field, and the form of the switching current due to a field changing linearly with time. Crystals of guanidine aluminium sulphate hexahydrate grown by both evaporation and rotary crystallization were cleft into seventeen portions between 0·15 and 0·8 mm thick, evaporated with gold electrodes 0·13 cm² in area and used in a dry atmosphere.

§ 2. Pulse Experiments

Voltage pulses of variable amplitude and alternate sign rising from a d.c. level adjusted to zero for all conditions were applied to the crystals. The rise time of these pulses was very short compared with the switching time of the crystals so that all measurements were made under conditions of steady state of voltage. The duration t_s and the maximum amplitude i_m of the resultant current flow were then observed as functions of the amplitude E of the switching pulses. At high field strengths both i_m and t_s were observed to fit the exponential type of law

$$i_m = i_0 \exp(-\alpha/E) \qquad \ldots\ldots(1)$$

(where α is a temperature dependent ' activation field '), which has been observed for $BaTiO_3$ by Merz (1954, 1956) and for guanidine aluminium sulphate hexahydrate by Holden *et al.* 1956. At low values of the field strength it was found that

$$i_m = \text{const.} \times E^2 \qquad \ldots\ldots(2)$$

$$t_s = \text{const.}/E^2. \qquad \ldots\ldots(3)$$

The behaviour characterized by equations (1) and (2) is shown in figures 1 and 2 respectively. This result was the same for many crystals, both with and without the permanent bias field which occurs in some crystals of guanidine aluminium sulphate hexahydrate (Holden *et al.* 1956).

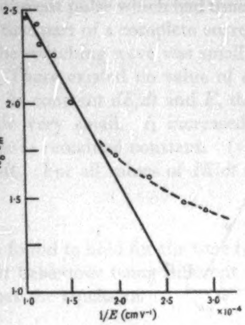


Figure 1.

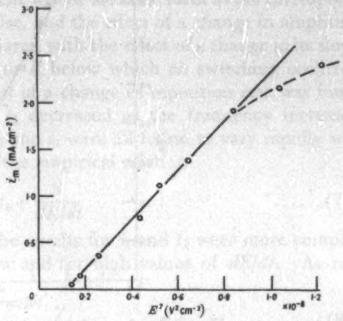


Figure 2.

Figure 1. Log of maximum switching current against $1/E$. The full line represents equation (1), the broken line the experimental results.

Figure 2. Switching current against E^2. The full line represents equation (2), the broken line the experimental results.

For a fixed value of the switching field strength the switching time was found to be proportional to the thickness of the crystal. The switching time at infinite field strength yields a maximum of 6×10^5 cm sec^{-1} for the velocity of domain wall motion. This figure is of the order of the velocity of sound in the material.

The values of the activation field α derived from the high field switching characteristics were about 5000 v cm^{-1} and were roughly inversely proportional to the crystal thickness.

§ 3. COERCIVITY EXPERIMENTS

If the switching current can be expressed in the form

$$i = F(P) \exp \{-\alpha/E(t)\} \qquad \ldots\ldots(4)$$

where $F(P)$ is a function of the polarization P, which changes slowly compared with the exponential term ; and if the switching time is inversely proportional to the frequency f, then Wieder (1957) has shown that a relation

$$\log f = \text{const.} - \alpha/E_c \qquad \ldots\ldots(5)$$

will hold for the coercivity E_c of $BaTiO_3$ crystals. Values of E_c taken as the value of the field strength when $P = 0$, were measured for guanidine aluminium sulphate hexahydrate crystals at frequencies up to 1 kc/s, at constant temperature and switching amplitude. Values of α found from a plot of $\log f$ against $1/E_c$ (figure 3), were between 15 000 and 17 000 v cm^{-1} for several crystals. These values are about three times larger than those found in other experiments, and a better approximation to the behaviour of the switching time is

$$t_s \propto 1/f^3. \qquad \ldots\ldots(6)$$

Since the effects of varying frequency, amplitude and rate of rise of field could not be separated in these sine wave variations of the switching field an investigation of the switching current due to a field increasing linearly with time was carried out.

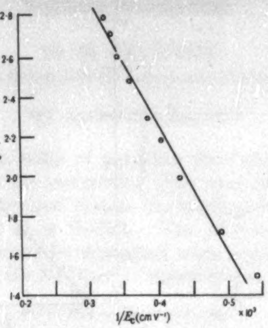


Figure 3. Log of switching frequency against $1/E_c$.

§ 4. LINEAR FIELD EXPERIMENTS

A saw-tooth pulse was generated with variable amplitude E of up to 500 volts, variable duration t of 200 μsec to 5 msec and repetition rates f of 15 c/s up to 1 kc/s. The base line of this saw-tooth was arranged to be at such a d.c. level V_0 that the crystal was biased permanently in one direction in the absence of a switching wave as shown in figure 5. The current i_1 and the times t_d and t_1 indicated in figure 4 were then measured as functions of dE/dt, E, f and pulse duration t as the field increased linearly down one side of the hysteresis loop.

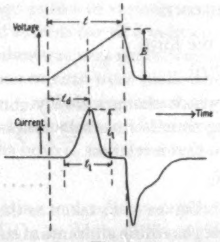


Figure 4.

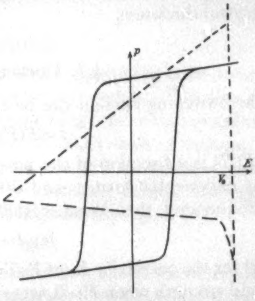


Figure 5.

Figure 4. Switching field and switching current against time.
Figure 5. The arrangement of the saw-tooth (broken line) with respect to the hysteresis loop.

The results of these experiments indicated that: (i) If the switching pulse was sufficiently long for polarization reversal to be completed then the size of the current pulse was independent of the length and amplitude of the switching pulse. (ii) If the switching pulse was not long enough for the polarization to reverse completely so that minor loops were being traversed, then that fraction of the current pulse which had time to occur was of the same form as the corresponding first part of a complete current pulse, and the effect of a change in amplitude of the switching wave was small compared with the effect of a change in its slope. (iii) There existed no value of dE/dt or E below which no switching occurred. (iv) At constant dE/dt and E, the effect of a change in repetition rate was found to be very small. i_1 increased and t_1 decreased as the frequency increased, while t_d remained constant. (v) t_d, t_1 and i_1 were all found to vary rapidly with dE/dt. For all values of dE/dt used, the empirical relation

$$t_d = t_0 + \frac{\beta}{dE/dt} \qquad \ldots\ldots(7)$$

was found to hold for the time t_d. The results for i_1 and t_1 were more complex, their behaviour being different for low and for high values of dE/dt. At high slopes the equations

$$t_1 = t_1 + \frac{\gamma}{dE/dt} \qquad \ldots\ldots(8)$$

$$\frac{1}{i_1} = \frac{1}{i_0} + \frac{\delta}{dE/dt} \qquad \ldots\ldots(9)$$

were found to hold. These are, in fact, good approximations to the exponentials

$$t_1 = t_1 \exp \frac{\gamma}{t_1\, dE/dt} \qquad \ldots\ldots(10)$$

$$i_1 = i_0 \exp \frac{-i_0\delta}{dE/dt} \qquad \ldots\ldots(11)$$

which is the result expected from the results of the pulse experiments at high E. The values of the activation field calculated from equations (10) and (11) agree closely with those found in the pulse experiments.

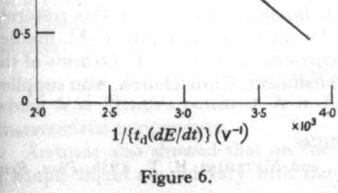


Figure 6.

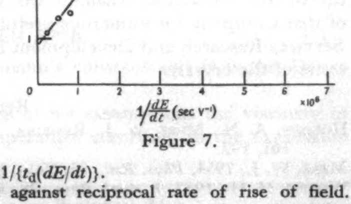


Figure 7.

Figure 6. Log peak switching current against $1/\{t_d(dE/dt)\}$.

Figure 7. Reciprocal peak switching current against reciprocal rate of rise of field. The full line represents equation (9), the broken line the experimental results.

However, these exponentials would not fit the results at low switching field strengths, the constants t_1 and i_0 in equations (10) and (11) appearing to be functions of dE/dt. t_1 was found to behave like t_d for small values of dE/dt as shown in figure 6.

The values of the constants t_0 and t_1 in equations (7) and (8) were found to be proportional to the crystal thickness and to represent a maximum velocity of domain wall motion of about $6 \times 10^2\,\mathrm{cm\,sec^{-1}}$ which is much lower than the values found in the pulse experiments. This probably means that equations (7), (8) and (9) are also not good approximations to the true behaviour of polarization reversal at very high dE/dt.

Since the rate of rise of field in a sine wave voltage variation is roughly proportional to its frequency it might be expected that a plot of $\log dE/dt$ against the reciprocal coercivity would show whether the coercivity is principally dependent upon the rate of rise of voltage or upon the repetition rate. Such a plot (figure 6) showed that

$$\log\frac{dE}{dt} = -\frac{C}{t_d\,dE/dt} + \text{const.} \qquad \ldots\ldots(12)$$

where C is a constant about two or three times larger than the value of the activation field found from the plots of $\log f$ against $1/E_c$. This is probably because the value of the field $t_d\,dE/dt$ is greater than E_c. Nevertheless, as this relation is of the same form as that found for E_c it suggests that the coercivity is principally dependent upon dE/dt and not f.

§ 5. CONCLUSIONS

(i) The polarization reversal process in guanidine aluminium sulphate hexahydrate results in peak currents i_m which can be expressed in the form

$$i_m = F(P)\exp(-\alpha/E) \qquad \ldots\ldots(13)$$

where the rate of change of the function $F(P)$ is significant only at very low and high rates of change of switching field. This is different from the behaviour of $BaTiO_3$ reported by Merz in which $F(P)$ can be regarded as constant compared with $\exp(\alpha/E)$. Nevertheless the problem of polarization reversal is essentially the same as in $BaTiO_3$ in that it appears to be principally a nucleation process.

(ii) The rate at which the polarization of a crystal reverses depends much more critically upon the rate of supply of energy than upon the total amount of energy.

ACKNOWLEDGMENTS

The author wishes to thank Dr. A. D. Booth for supervising this research, the British Tabulating Machine Co. Ltd. for their assistance, Mr. E. M. Bradley of that Company for numerous useful discussions, and Mr. C. F. Oxbrow of the Services Research and Development Establishment, Christchurch, who supplied some of the crystals.

REFERENCES

HOLDEN, A. N., MERZ, W. J., REMEIKA, J. P., and MATTHIAS, B. T., 1956, *Phys. Rev.*, **101**, 192.
MERZ, W. J., 1954, *Phys. Rev.*, **95**, 690 ; 1956, *J. Appl. Phys.*, **27**, 938.
WIEDER, H. H., 1957, *J. Appl. Phys.*, **28**, 367.

The Viscosity of Liquid Indium and Liquid Tin

BY M. F. CULPIN†
Davy Faraday Laboratory, Royal Institution, London

M.S. received 24th April 1957, *and in final form* 21*st August* 1957

Abstract. Measurements of the viscosity of liquid indium and liquid tin have been made, using the oscillating sphere viscometer. A preliminary set of measurements on bromine was made. The viscosity was determined over a range of temperature in each case, and the results compared with Andrade's theory of the viscosity of liquids. A comparison is also made between the many sets of values of the viscosity of tin already published.

§ 1. INTRODUCTION

A THEORY of the liquid state may be considered as entirely satisfactory only if it is possible to predict from it the value of the viscosity, amongst other things, in terms of the properties of the molecular components of the liquid. From the simple description of liquid structure proposed by Andrade (1934) the following expression for the viscosity at the melting point is derived:

$$\eta_{\mathrm{m}} = \alpha \frac{\nu m}{\sigma} \qquad \ldots\ldots(1)$$

where η_{m} is the viscosity, m the mass of a molecule, ν the vibrational frequency of the molecules, σ their mean spacing, and α a factor, of the order of 4/3, which is proportional to the probability of a molecular interaction leading to complete momentum sharing.

Two possible sources from which ν can be obtained are the Lindemann melting point formula, used in conjunction with the simple Einstein theory of specific heats of solids, and the characteristic temperature as defined in Debye's theory of the specific heats of solids. In the first case we may derive the expression

$$\eta_{\mathrm{m}} = K(AT_{\mathrm{m}})^{1/2}(A/\rho)^{-2/3} \qquad \ldots\ldots(2)$$

where A is the atomic weight, T_{m} the absolute temperature of melting, and ρ the density. K is equal to $\alpha C/\boldsymbol{N}^{2/3}$ where $C=(\sqrt{2}/\pi n)\boldsymbol{R}^{1/2}$, $\boldsymbol{R}$ is the gas constant for a gramme atom, n is the ratio of the shortest distance between the spheres of influence of two atoms to the distance between their centres, and $\boldsymbol{N}$ is Avogadro's number. The latter source gives

$$\eta_{\mathrm{m}} = \alpha(A^2\rho)^{1/3}\boldsymbol{k}\Theta/\boldsymbol{N}^{2/3}\boldsymbol{h} \qquad \ldots\ldots(3)$$

where $\boldsymbol{h}$ is Planck's constant, $\boldsymbol{k}$ is Boltzmann's constant and Θ is the Debye characteristic temperature.

Andrade also showed that on the basis of his assumptions the viscosity of a simple liquid should vary with the temperature according to the expression

$$\eta/\rho^{1/3} = b\exp(c\rho/T) \qquad \ldots\ldots(4)$$

---

† Now at British Nylon Spinners, Pontypool, Mon.

where η is the viscosity, ρ the density, T the absolute temperature, and b, c are constants.

Among the liquids most likely to conform to Andrade's model are the metals. It would therefore be of interest to compare equations (1) and (4) with reliable experimental data for metals. Although there has been an increase recently in the volume of published work on the viscosity of liquid metals, there remain many inconsistencies between results obtained by different workers for the same metal and a need for work on further metals. The work described here deals with tin, for which a large number of more or less discordant results have been published, and with indium, for which there are no results already published. A set of measurements of the viscosity of bromine is also reported, which, by its agreement with two previously published sets of figures, was assumed to show that the present experimental technique was satisfactory.

§ 2. GENERAL EXPERIMENTAL METHOD

The method used for measuring the viscosity was that of observing the damping of the oscillations of a hollow sphere filled with the liquid and supported by a bifilar suspension (Andrade and Chiong 1936). From measurements of the damping of the motion and the period of oscillation both when the metal is solid and when it is liquid, the viscosity may be calculated, thus:

$$\eta = \frac{\pi\rho a^2 R^2 p}{16\tau_2}\{1-(1-\mu)^{1/2}\}^2 \qquad \ldots\ldots(5)$$

where $p = 1 + (\eta'\tau_2/\pi\rho R^2)$; $\mu = 3M\tau_2/\pi^2 R^5 a^2 \rho p$;

$$a = 1 - (\Delta_2/4\pi); \qquad M = \frac{I\Delta_2}{\tau^2}\left\{\frac{\tau_2^2}{\tau_0^2}+1\right\} - \frac{I+I'}{\tau_1}\Delta_1\left\{\frac{\tau_1^2}{\tau_0^2}+1\right\}$$

ρ is the density of the liquid; R the radius of the sphere; Δ_1, τ_1 the logarithmic decrement and period of oscillation when the metal is solid; Δ_2, τ_2 the same when the metal is liquid; I the moment of inertia of the system apart from the metal; I' the moment of inertia of the metal when solid; and η' an approximate value of η. The method of successive approximations may be used to obtain η to any desired accuracy. If Δ_1 is not greater than a few per cent of Δ_2, then the expression for μ may be simplified to $I(\Delta_2-\Delta_1)(\tau_2^2/\tau_0^2+1)/\pi^2 R^5 a^2 \rho p$ with sufficient accuracy.

The experimental arrangement is shown in figure 1. In practice it is necessary to increase the moment of inertia of the sphere A by means of a cross-bar B, which may carry adjustable masses C to enable the moment of inertia to be measured. The system should also carry a mirror D for use in recording photographically the amplitude of the oscillations, and another E for use with a simple photocell arrangement for counting the oscillations. The suspension is a single wire F passing round a small pulley wheel G on the viscometer. The ends of the wire are attached to a head H which may be rotated through a few degrees about a vertical axis to impart a rotational movement, also about a vertical axis, to the sphere. The rotation of the head may conveniently be made by a simple device employing two electromagnets J, arranged to pull a lever K in opposite directions. Operating one magnet rotates the head quickly through a fixed angle and imparts a certain amplitude of oscillation to the sphere. When this has been reduced to a low value, the other magnet may be operated to give an impulse in the opposite direction, but of the same magnitude as before, thus starting another set of oscillations.

It is necessary to enclose the entire viscometer and suspension within an evacuated chamber to reduce the damping of the system due to air drag. To eliminate the effects of stray vibrations it is convenient to mount the whole apparatus upon girders let into a solid wall. Levelling screws allow the system to be adjusted to bring the axis of the head accurately vertical.

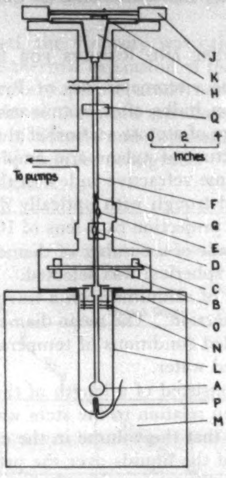


Figure 1. General arrangement of the apparatus.

The temperature of the material in the sphere was controlled by immersing the Pyrex glass tube L, which constitutes the lower part of the vacuum chamber, in a water bath or low temperature furnace M. Radiation shields N complete the uniform temperature enclosure of the sphere, whilst an asbestos plate O serves to keep heat from the inertia system and suspension. A thermocouple in an inset tube P was used to measure the temperature of the metal (Chiong 1936).

In setting up the viscometer it was useful to be able to observe the suspension through the window Q.

The moment of inertia was calculated from measurements of the period of oscillation of the system for different settings of the adjustable masses on the inertia bar. The moment of inertia of the masses being known from their geometry, the sets of measurements could be taken in pairs and the unknown moment of inertia calculated.

Calculations of the logarithmic decrement were made from records on a strip of photographic paper of the successive amplitudes of oscillation. An image of a stationary slit was thrown by a lens, via a mirror on the viscometer, on to the photographic paper. When the viscometer came to rest at the end of each swing, the image of the slit was recorded, together with part of an engraved Perspex scale placed in front of the paper. When the damping was small it was convenient to record every fifth or twenty-fifth swing, in order to measure over a large range of amplitude without completely blackening the paper. To do this a shutter was placed in the optical system, and was operated by a scaling device driven by a lamp, mirror and photocell system, the mirror being on the viscometer.

Details of the techniques of measuring the temperature of the metal in the sphere, the period of oscillation, the moment of inertia, and the logarithmic decrement are given by Andrade and Chiong (1936), Chiong (1936) and Dobbs (1952).

§ 3. Experimental Procedure and Results for Bromine, Indium and Tin

The sphere used for the viscometer was of Pyrex glass, and was selected from a batch of hand-blown bulbs after examination by a projection method. To obtain an enlarged image of a cross section of the inner wall, the bulbs were immersed in a suitable mixture of xylene and amyl acetate, shown by a Becke line method to have the same refractive index as the glass. The mixture was contained in a parallel-sided trough with optically flat sides. Illumination was by a Point-o-Lite lamp and projection by a lens of 10 cm focal length. On each bulb measurements were made of a number of diameters of each of several cross sections. The most nearly spherical was selected. Its equatorial diameter was then measured at a number of azimuths with a travelling microscope, using the same liquid mixture for immersion. The mean diameter of the bulb was obtained by weighing, under controlled conditions of temperature, the bulb when empty and when filled with distilled water.

The stem of the bulb consisted of a length of thick walled capillary tubing. The symmetry of the bulb in relation to the stem was taken into account in the selection. It was estimated that the volume in the capillary would be sufficient to allow for the expansion of the liquids over the proposed temperature ranges, but that the radius of the capillary was so small that any oscillation of the liquid in the stem would contribute a negligible amount to the damping of the viscometer.

To ensure that the experimental technique was being carried out correctly, and also to serve as a check on the dimensions of the sphere, a liquid was sought for which reliable viscosity data were available. It was required that its viscosity and density should be as close as possible to those of the metals which were to be tested. The choice of bromine was made in view of the concordance between the results of Thorpe and Rodger (1894) and of Steacie and Johnson (1925) for the viscosity; and of Thorpe (1880), Andrews (1907) and Trautz and Triebel (1927) for the density.

The viscometer was first set up with the bulb empty, and the 'residual damping' and moment of inertia measured. A vacuum pressure of about 10^{-5} mm of mercury was maintained during these measurements, and also during the subsequent work with the sphere full of liquid. The viscometer was then dismantled and a funnel attached to the upper end of the stem. By a careful warming and cooling technique a quantity of Analar bromine was introduced such that it just filled the bulb when the temperature was just above the melting point. To remove the air from above the bromine, the stem was next sealed to a vacuum system, the bromine frozen, and the stem evacuated and sealed off at its upper end.

The viscometer was re-assembled, and measurements of the damping and period of oscillation made over a range of temperatures above the melting point. Agreement to within one part in a thousand between values measured at intervals of a quarter of an hour was taken as evidence that a steady temperature had been reached in the sphere. A water bath was used to surround the lower end of the vacuum enclosure, the temperature being controlled by a mercury and toluene regulator and Sunvic hot wire vacuum switch. Measurements below room temperature were made using aniline, water and glacial acetic acid as cryostat materials.

At each temperature used, the viscosity was calculated as described in §2 Allowance was made for the thermal expansion of the sphere, whilst the density of bromine was estimated from a graph prepared from the three sources mentioned above. In figure 2 the results of this work are shown and compared with those of Thorpe and Rodger and of Steacie and Johnson. In this figure viscosity is plotted against temperature in degrees centigrade. The agreement between the sets of results is seen to be close.

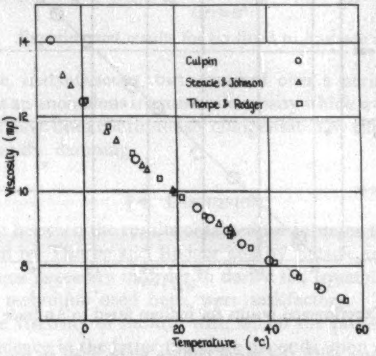


Figure 2. Viscosity of bromine.

After the measurements on bromine, the sphere was emptied and cleaned. It was then filled with indium. Owing to the very low vapour pressure of this metal, the sphere could not be filled by distillation. The stem was therefore attached to a funnel which was part of a vacuum system, and into which a suitable quantity of the metal was placed. After melting the metal was allowed to de-gas. Some of it was then forced into the sphere by admitting argon to the system. Although the pieces of metal in the funnel had been carefully scraped, some scum appeared on the surface at melting. This however remained in the funnel after the sphere had been filled. Finally, the system was again evacuated, and the stem sealed off and detached at the upper end.

The viscometer was set up in the vacuum enclosure as used for bromine, but with the lower part surrounded by a low temperature furnace. This was wound non-inductively, but as an additional precaution against stray magnetic fields a thick iron tube was placed between the furnace and the viscometer. The temperature was controlled by a thermocouple used in conjunction with a slide wire potentiometer, mirror galvanometer, lamp and photocell. The latter operated a Sunvic switch through a suitable circuit.

A set of measurements was made of the damping of the viscometer when filled with liquid indium. To calculate the viscosity, the density was taken from a graph prepared from figures given by Gamertsfelder (1941) and by Williams and Miller (1950). The results are given in table 1. Figure 3 shows the results in the form of a plot of $\log_{10}(\eta/\rho^{1/3})$ against ρ/T. The linear relationship is in agreement with equation (2). Extrapolation to the melting point gives for η_m the value 19·4 mP, whilst the best values of the constants b and c in equation (2) were found by the method of least squares to be $2{\cdot}44\times10^{-3}$ and 86·5 respectively.

Table 1. The Viscosity and Density of Indium

Temperature (°c)	164·6	174·8	190·6	215·6	230·8	255·8	280·5	335·8
Density (g cm^{-3})	7·025	7·017	7·003	6·982	6·968	6·947	6·926	6·883
Viscosity (mP)	18·85	18·11	17·34	16·09	15·49	14·38	13·78	12·42

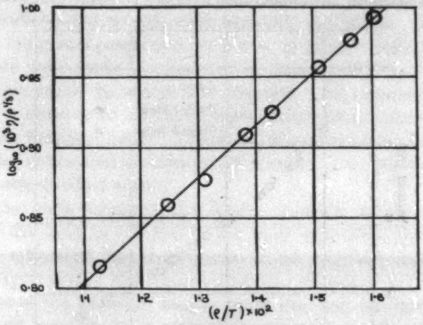


Figure 3. Experimental results for indium fitted to Andrade's equation.

Similar techniques with the same sphere were used to measure the viscosity of tin. In this case, figures for the density were taken from a graph based upon values by Plüss (1915), Day, Sosman and Hostetter (1915), Sauerwald and Töppler (1926) and Matsuyama (1929). The results of this work, shown in table 2 and figure 4, yield for η_m the value 21·4 mP, and for the constants b and c the values $2{\cdot}30\times10^{-3}$ and 114 respectively.

In view of the published results, which will be discussed below, showing a sharp increase in the viscosity near the melting point, a further experiment was performed with tin. The temperature was raised to a steady value a few degrees above the melting point, and then allowed to fall at the rate of one degree in several minutes. The amplitude of oscillation was recorded continuously until the freezing point was past. It may be assumed that the temperature of the metal during this time was in a quasi-steady state, and that any anomalous increase in the viscosity near the melting point would have caused a marked increase in the damping at some time in the middle of such a run. No increase was observed. Instead, the *slight* increase, due to the fall in temperature, suddenly gave way to a drop, indicating that solidification had occurred. The decrease

Table 2. The Viscosity and Density of Tin

Temperature (°c)	238·2	253·3	274·1	331·5
Density (g cm^{-3})	6·974	6·962	6·946	6·900
Viscosity (mP)	20·92	19·98	18·76	16·21

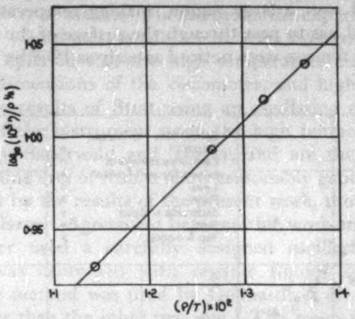


Figure 4. Experimental results for tin fitted to Andrade's equation.

was not, of course, instantaneous, but occurred over a period of minutes. It seems unlikely that an anomalous increase in viscosity which would have increased the damping could have been continuously compensated by effects of solidification tending to decrease the damping.

§ 4. DISCUSSION

The agreement between the results obtained for bromine in the present work, and those obtained by Thorpe and Rodger and by Steacie and Johnson suggest that the assumptions necessary in order to derive the equations given in § 2, and the experimental technique used here, were satisfactory. There has been no other work on the viscosity of indium with which the present results could be compared. Confidence in the latter therefore depends upon the results obtained for bromine. In the case of tin, however, there are many published sets of figures which it may be useful to consider in some detail.

The values obtained by different workers for the viscosity of tin are shown in figure 5. In order to compare the values of η_m obtained by extrapolation to the melting point, a plot of $\log \eta$ against $1/T$ is shown in figure 6. Some of the sets in figure 5 have been omitted from figure 6. The early results of Arpi, obtained by a capillary method, are obviously inaccurate in comparison with all later work. Those of Bootov, Priss and Shvidkovski (1951) show the viscosity to be considerably lower than was found by all others except Arpi (1914), and in addition show an irregularity in the viscosity against temperature curve at about 100°c above the melting point, which has not been observed by other workers. They used an oscillating hollow cylinder viscometer, but it is felt that their elaborate theoretical treatment of the method has not been accompanied by an adequate experimental technique.

The results of Yao and Kondic (1952) show considerable scatter and at temperatures approaching the melting point show a very marked increase in the viscosity. While the authors claim that this effect is real, and also claim to have shown it in other metals, it has not been observed by any other workers except Jones and Bartlett (1952) who used a similar type of viscometer. The observed increase is therefore probably due to some feature of the experimental arrangement, which was of the oscillating solid of revolution type. In this the experimental liquid is distributed over a large region, making temperature control difficult, and a supporting rod has to pass through the surface of the liquid. Undesirable effects due to surface tension were noticed as early as 1908 by Fawsitt with systems of this kind.

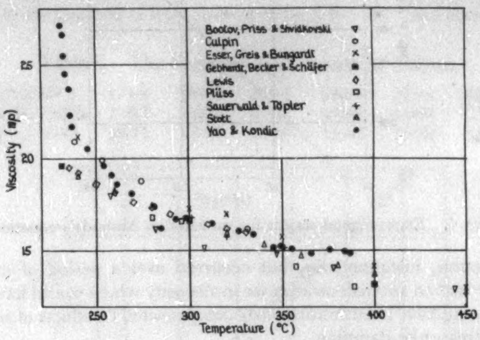


Figure 5. Viscosity of tin, as determined by a number of workers.

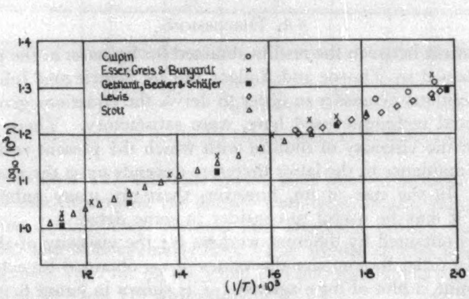


Figure 6. Relationship between log η and $1/T$ for some of the sets of figures of figure 5.

Sauerwald and Töpler obtained results over a range of temperature above 300°c or seventy degrees above the melting point, and these have not therefore been used in figure 6. Although their capillary viscometer appears to have been as well designed as is possible with such an instrument, the hazards of partial blockage of the capillary, of surface tension effects, and of non-uniformity of the temperature conditions were nevertheless present. The same arguments apply to the results of Plüss, which again cover a range of temperatures well above the melting point.

Considering now the results plotted in figure 6, and extrapolating to the melting point, the following values of η_m are obtained: Esser, Greis and

Bungardt (1934), 21·58; Stott (1933), 20·00; Lewis (1936), 19·50; Gebhardt, Becker and Schafer (1952), 19·54; the present work, 21·38. The mean of these is 20·40. The results obtained by Lewis (1936) depart noticeably from a straight line when plotted as $\log\eta$ against $1/T$ and in this respect disagree with most of the work on the viscosity of metals. Furthermore, the experimental arrangement of Lewis, a capillary viscometer, is subject to even stronger criticism than that of Sauerwald and Töpler in that the high temperature enclosure was smaller compared with the dimensions of the viscometer, and high vacuum conditions were not used. The results of Stott using an oscillating disc method depend for the calibration of the instrument upon the high temperature values of the viscosity obtained by Sauerwald and Töpler, and are therefore not absolute.

The three remaining sets of results show reasonably good fit to straight lines. The best fit is shown by the results of the present work, though the temperature range covered is not large. Agreement between this work and that of Esser *et al.* is good. The latter used a carefully designed oscillating hollow cylinder viscometer which was calibrated with organic liquids of accurately known viscosity. A similar method was used by Gebhardt *et al.* though their results are consistently lower than the other two sets. The mean value of η_m obtained from all three sets of figures is 20·8 mP, which is perhaps the best available value of the melting point viscosity of tin. The agreement between two of the three sets suggests that a slightly higher weighted mean should be taken.

Taking the results of the present work in relation to the expressions quoted from Andrade's theory in § 1 we find that the melting point viscosities of indium and tin correspond to values of α of 1·30 and 1·33 respectively, which are very close to the roughly estimated theoretical value of 4/3.

A figure for the Debye characteristic temperature could only be found for one of the metals examined in this work, viz. tin. It is quoted by James (1948), and gives for α the value of 1·23. In view of the simplifications upon which the Andrade theory is based, the agreement between the two values of α for tin may be considered as very good, and to show the importance of the frequency factor in the mechanism of momentum transfer in liquid metals.

ACKNOWLEDGMENTS

I wish to record my thanks to Professor E. N. da C. Andrade for his supervision of this work, and also to Professor R. King for his help and advice. To the Department of Scientific and Industrial Research I am indebted for a maintenance grant, and to the Managers of the Royal Institution for laboratory facilities.

REFERENCES

ANDRADE, E. N. DA C., 1934, *Phil. Mag.*, **17**, 497, 698.
ANDRADE, E. N. DA C., and CHIONG, Y. S., 1936, *Proc. Phys. Soc.*, **48**, 247.
ANDRADE, E. N. DA C., and DOBBS, E. R., 1952, *Proc. Roy. Soc.* A, **211**, 12.
ANDREWS, L. W., 1907, *J. Amer. Chem. Soc.*, **26**, 688.
ARPI, R., 1914, *Z. Metallk.*, **5**, 142.
BOOTOV, A. M., PRISS, L. S., and SHVIDKOVSKI, E. G., 1951, *J. Tech. Phys.*, **21**, No. 11.
CHIONG, Y. S., 1936, *Proc. Roy. Soc.* A, **157**, 264.
DAY, A. L., SOSMAN, R. B., and HOSTETTER, J. C., 1915, *Neues Jb. Min. Blg. Bd.*, **40**, 119.
ESSER, H., GREIS, F., and BUNGARDT, W., 1934, *Arch. Eisenhüttenw.*, **7**, 385.
FAWSITT, C. E., 1908, *Proc. Chem. Soc.*, **24**, 146.

GAMERTSFELDER, C., 1941, *J. Chem. Phys.*, **9**, 455.
GEBHARDT, E., BECKER, M., and SCHÄFER, S., 1952, *Z. Metallk.*, **43**, 292.
JAMES, R. W., 1948, *The Crystalline State*, Vol. II, p. 221 (London : Bell).
JONES, W. R. D., and BARTLETT, W. L., 1952, *J. Inst. Met.*, **81**, 145.
LEWIS, A. J., 1936, *Proc. Phys. Soc.*, **48**, 102.
MATSUYAMA, T., 1929, *Sci. Rep. Tôhoku Univ.*, **18**, 20.
PLÜSS, M., 1915, *Z. anorg. Chem.*, **93**, 1.
SAUERWALD, F., and TÖPLER, K., 1926, *Z. anorg. Chem.*, **157**, 117.
STEACIE, E. W. R., and JOHNSON, F. M. G., 1925, *J. Amer. Chem. Soc.*, **47**, 754.
STOTT, V. H., 1933, *Proc. Phys. Soc.*, **45**, 530.
THORPE, T. E., 1880, *J. Chem. Soc.*, **37**, 141, 327.
THORPE, T. E., and RODGER, J. W., 1894, *Phil. Trans. Roy. Soc.*, A, **185**, 397.
TRAUTZ, M., and TRIEBEL, E., 1927, *J. prak. Chem.*, **116**, 205.
WILLIAMS, D. D., and MILLER, R. R., 1950, *J. Amer. Chem. Soc.*, **72**, 3821.
YAO, T. P., and KONDIC, V., 1952, *J. Inst. Met.*, **81**, 17.

The Viscosity of Liquid Magnesium and Liquid Calcium

By M. F. CULPIN†
Davy Faraday Laboratory, Royal Institution, London

M.S. received 24th April 1957, *and in final form* 21*st August* 1957

Abstract. In a previous paper measurements of the viscosity of indium and tin are described. The experimental technique described there, viz. the method of the oscillating hollow sphere, has been modified to enable measurements of the viscosity of liquid magnesium and calcium to be made.

§ 1. Introduction

The oscillating sphere viscometer (Andrade and Chiong 1936) has a number of advantages over other methods of measuring the viscosity of liquid metals. This is especially so in the case of highly reactive metals, when it provides almost the only satisfactory means of viscosity measurements (Chiong 1936, Andrade and Dobbs 1952). The last two papers describe work on the series of alkali metals. From a theoretical point of view it would be useful to have experimentally determined values of the viscosity of all the metals in another group of the periodic table. At the time when this work was begun there were no figures available for any of the alkaline earth metals, so the possibility was considered of adapting the oscillating sphere viscometer for use with these.

§ 2. Apparatus

The general disposition of the apparatus used for this work was similar to that already described (Culpin 1957). Differences in detail arose from the requirements that the temperature of the metal should be raised to about 1000°c, and that the material of the sphere should be inert at such temperatures to the metals to be used. Leaving out consideration of beryllium and radium, for which there would be special experimental difficulties, a material was sought which is not attacked by magnesium, calcium, strontium or barium in the liquid state. Owing to lack of information on the resistance of different materials to attack by these metals, it was necessary first to consider those from which accurately spherical bulbs could be made. The next consideration was whether the bulbs would stand up to the high temperatures. Finally, it was necessary to test for chemical attack by the liquid metals.

It seemed possible that graphite, fused alumina or fired steatite might be suitable. Of these graphite was observed to suffer no attack by magnesium up to 800°c; and while calcium was found to adhere firmly to graphite after heating to 900°c, there was no evidence of carbide formation. Barium was found to attack graphite vigorously above the melting point. No tests were made with strontium.

In the case of fused alumina, no attack by barium was observed, and a number of specially prepared alumina bulbs were examined for sphericity. This was done by taking a series of radiographs with each bulb. Unfortunately, none proved to be suitable for use in the viscometer.

---

† Now at British Nylon Spinners, Pontypool, Mon.

Barium was found not to attack fired steatite, at least up to 900°c, and preliminary tests on machining a hollow sphere from soapstone, and carefully firing it, gave encouraging results. However, the work on magnesium and calcium took longer than was anticipated, and time allowed no further progress with the design of a steatite viscometer.

The viscosities of magnesium and calcium were therefore determined using a sphere constructed of graphite, of the fine quality specified by the Morgan Crucible Co. as CY 9062.

2.1. *The Graphite Sphere*

This part of the viscometer is shown in figure 1. The hollow sphere A is supported by a stem B through which a small hole is bored. Above the stem a chamber C is fitted to allow for the expansion of the liquid metal in the sphere during the viscosity measurements. Rotational motion of the metal in the chamber is prevented by a vane, shown in section and attached to the plug D. A small hole in the wall of the chamber C allows air to escape from the viscometer when it is placed in an evacuated enclosure. By means of an alumina pin F the chamber C is supported inside a close-fitting collar E which, with the graphite rod G, joins the lower parts of the viscometer to the brass inertia system, with its mirrors, as used in the previous work. The same suspension was also used.

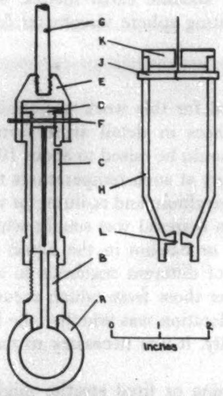


Figure 1. The graphite viscometer bulb and attachments.

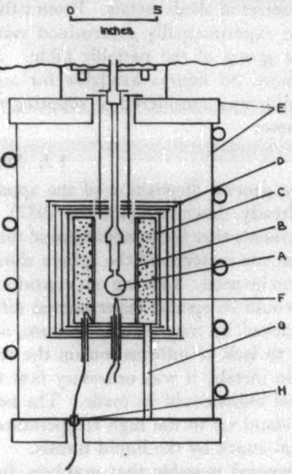


Figure 2. The high temperature vacuum enclosure.

The diameter of the sphere was measured indirectly. It was possible, using the lathe tool which was specially designed for boring out the hollow sphere, to make a cylindrical hole of the same diameter as the sphere, in another block of graphite. Measurements of the diameter of this hole could be made with an inside micrometer, and since precautions had been taken to ensure that the bulb was accurately spherical and not ellipsoidal, these measurements could be assumed to give the diameter of the bulb.

For filling the sphere with metal, a funnel H was screwed to the stem B in place of the chamber C. The graphite system could then be supported inside a suitable enclosure by means of an alumina pin J, above which is a lid K to reduce evaporation of the metal in the funnel. A quartz glass tube, closed at one end, and supported inside the vertical tube of an electric furnace, constituted the enclosure used for filling the sphere. It was also used for the tests for chemical attack, and for measuring the density of calcium.

To fill the sphere, pieces of metal of convenient size were scraped as clean as possible and placed in the funnel, which was suspended with the sphere inside the quartz glass tube. The tube was then evacuated, the temperature raised to just below the melting point of the metal, and time allowed for the metal to de-gas. The graphite system and quartz glass tube had previously been de-gassed at 1000°c and returned to atmospheric pressure under specially cleaned argon. This supply of argon was also used to force metal into the sphere after melting, and to return the graphite system to atmospheric pressure before cooling. In this way, oxygen, nitrogen and water vapour could be kept from the metal. A similar melting procedure was used in the tests for chemical attack, the material under test being in the form of a small crucible which was suspended within the quartz glass tube.

After removal from the tube, the funnel could be unscrewed from the stem, in the case of magnesium, breaking the thread of metal between them. In the case of calcium the funnel had to be detached by cutting through the stem just below the funnel. In both cases a check was made by weighing to ensure that sufficient metal was contained in the sphere to fill it at the melting point.

2.2. *The High Temperature Vacuum Enclosure*

It is necessary to use the oscillating sphere viscometer under high vacuum conditions to reduce the damping of the motion of the viscometer due to causes other than the viscosity of the liquid contained in it. For work on magnesium and calcium an enclosure was constructed of which the upper parts have been described previously (Culpin 1957). These will not be discussed again. The lower parts were designed to house an electric furnace for use up to 1000°c and to prevent heating of the upper parts. Figure 2 shows the general arrangement.

The sphere A is suspended within the refractory tube B upon which there is a doubly wound non-inductive winding of nichrome wire. Electromagnetic damping of the viscometer due to imperfect winding was, however, observed and will be discussed below. Surrounding the winding is a layer of alumina powder, contained in a steel shell C. Thin steel radiation screens D reduce the heat lost to the water-cooled cylinder and end plates E. The temperature of the metal in the sphere is measured with a thermocouple F. The calibration of this thermocouple with reference to a second thermocouple placed inside the sphere is described below. Three steel rods, one of which is shown at G, support the furnace whilst the leads to the winding and to the thermocouple are taken through ceramic bushes. One of these is shown at H.

Two power supplies were constructed for the furnace, one for direct current and one for alternating current. Temperature control could be obtained with either by means of the thermocouple F and a photoelectric amplifier, which was arranged to switch the current between high and low values. The alternating supply was most convenient to use, but in view of the presence of electromagnetic damping it was necessary to observe whether it was influenced by the kind of current in the furnace.

For the sake of clarity, the constructional details of the high temperature enclosure have been omitted from figure 2. The parts which had to be separated in the course of the tests were all marked in such a way that the enclosure could be re-assembled with the sphere always in the same place relative to the furnace.

The weight of the entire apparatus was taken by girders set into a solid wall below the furnace enclosure, whilst the upper parts were steadied by other girders to reduce the influence of stray vibrations.

§ 3. EXPERIMENTAL METHODS FOR MAGNESIUM

The magnesium for this work was supplied by the National Physical Laboratory, and was in the form of a block from which pieces could be cut. Analysis by Messrs. Johnson, Matthey showed it to contain 0·02% of nitrogen, with very small quantities of calcium, copper, silicon and sodium as the major metallic impurities.

For the measurements of the viscosity the graphite viscometer was filled with magnesium by the method described in §2. It was then set up in the vacuum enclosure, and the moment of inertia and period of oscillation measured, with the metal in the solid state, as described elsewhere (Culpin 1957).

Owing to the presence of a small magnetic field due to the electric furnace, the problem of the damping of the motion of the viscometer by factors other than the viscosity of the liquid had to be considered in some detail. The motion may be damped by: (i) the drag due to liquid in the sphere; (ii) the eddy current damping due to the magnetic field; (iii) drag due to residual gases around the viscometer; (iv) loss of energy due to internal friction in the suspension wire. It is convenient for the purpose of calculation if the damping due to factors (ii), (iii) and (iv) together is no more than a few per cent of that due to (i), when it is only necessary to subtract the residual damping from that observed when the bulb is filled with liquid. If this is not the case, as in the work described here, the full procedure of calculation given in the last reference must be used.

This method of correction strictly demands a measure of the damping due to causes (ii), (iii) and (iv) while the metal is in the liquid state. In practice this cannot be achieved so it will be assumed that the measurements now discussed give it with sufficient accuracy. The assumptions made are that the damping due to (iii) is independent of temperature, and that that due to (ii) is dependent on the furnace current only and not on the temperature, and is approximately the same for the liquid metal as for the solid.

With the viscometer in position in the enclosure, the value of the current necessary to maintain the furnace at each of a number of temperatures was determined, and a graph plotted of temperature against current. Then, starting with a cold furnace each time, damping measurements were made with the furnace current set at a number of different values. These included some which would, under steady conditions, maintain the temperature above the melting points of the metals to be used, but in these preliminary tests the temperature was never allowed to rise to the melting point of magnesium. The values of the damping thus observed were plotted against furnace current and are shown in figure 3 (upper curve). The linear relationship is strong support for the assumption that the damping due to causes other than the viscosity of the metal was independent of temperature. Additional evidence is shown in the lower curve, obtained by allowing the furnace temperature to become steady at some point below the melting point of magnesium, after which the current was switched off and a damping measurement made. The thermal inertia of the system was such that in the course of the damping measurement the temperature fell by only a few degrees. In the lower curve of figure 3 the damping is plotted against the current which would have been required to maintain the steady temperature of the furnace before it began to cool. The horizontal line on which the points lie shows that the residual damping is independent of temperature, and that the value of the residual damping required to correct subsequent viscosity measurements could be obtained from the upper curve, according to the current used during those measurements.

The residual damping having been examined, the damping of the motion of the viscometer and its period of oscillation were measured at a number of temperatures at which the magnesium was in the liquid state. To ensure that no measurements were made with the sphere partly empty due to metal having been left in the expansion chamber, all measurements except those near the melting point of the metal were made at temperatures in a rising sequence.

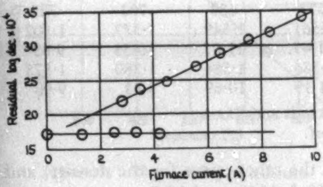


Figure 3. The residual damping of the motion of the viscometer as a function of furnace current.

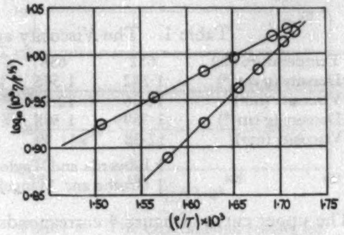


Figure 4. The experimental results for magnesium fitted to Andrade's equation.

To investigate changes of viscosity with temperature very close to the melting point, measurements of the damping were made whilst the viscometer cooled slowly through the freezing point of the metal. These showed that there was no anomalous increase of the viscosity near the freezing point, and also that the value of the melting point as given by the thermocouple below the sphere during these measurements was several degrees lower than the published figure for the pure metal. When all the damping measurements had been made, therefore, this thermocouple was calibrated in terms of the temperature indicated by a second thermocouple placed actually inside the sphere.

The system was allowed to cool. It was then dismantled, and a hole was drilled into the sphere to take an alumina tube for the second thermocouple, whose junction was to lie at the centre of the sphere (see Chiong 1936). The system was completely re-assembled, the enclosure evacuated, and the furnace raised to a number of steady temperatures. In each case the temperature of both thermocouples was observed. From the calibration curve which was plotted from these data, it was found that the temperature of the melting point of the magnesium was in fact close to the published figure. As an additional check, heating and cooling curves were taken with the apparatus as it now stood. These gave the melting point as 649·5°c. After this check, the calibration was continued over a range of temperatures to include those necessary for the work on calcium.

In order to calculate the viscosity of the metal, the value of the density is required. Values were at first taken from a curve prepared from figures given by Arndt and Ploetz (1927) and Edwards and Taylor (1923). It was later found that the measurements of Grothe and Mangelsdorf (1937) have a curve markedly different from the other. Figures quoted by Gebhardt, Becker and Trägner (1955) as being taken from Kubaschewski and Hörnle (1951), but which were in fact obtained by the authors themselves, agree very well with Grothe and Mangelsdorf's values. The difference between the last mentioned two sets of figures is not sufficient to justify calculating the viscosity using both sets. Therefore, figures for the viscosity based upon the first mentioned sets of figures for the density are given in table 1. The results are plotted as $\log_{10}(\eta/\rho^{1/3})$ against ρ/T in figure 4, and show good agreement with Andrade's formula for the variation of viscosity with temperature. The lower curve of figure 4, corresponding to Grothe and Mangelsdorf's figures for the density, and plotted from table 1, gives for b, c and η_{m} the values $2{\cdot}45 \times 10^{-4}$, 2190 and 12·3 mP respectively.

Table 1. The Viscosity and Density of Magnesium

Temperature (°c)	652	658	672	683	703	725
†Density (g cm^{-3})	1·582	1·576	1·561	1·549	1·527	1·502
†Viscosity (mP)	12·36	12·13	11·47	11·07	10·35	9·63
‡Density (g cm^{-3})	1·589	1·588	1·586	1·584	1·580	1·575
‡Viscosity (mP)	12·28	11·99	11·19	10·69	9·83	9·00

† Edwards and Taylor, Arndt and Ploetz.
‡ Grothe and Mangelsdorf.

The upper curve of figure 4 corresponds to the other curve for the density, and gives for these quantities the values $1{\cdot}50 \times 10^{-3}$, 1140, and 12·5 mP respectively.

§ 4. EXPERIMENTAL METHODS FOR CALCIUM

Calcium for this work was supplied by the Atomic Energy Research Establishment in the form of pellets. For cleaning purposes these were cast into block form in a graphite crucible, as described above. The graphite was turned away from the calcium, which could then be cleaned, when required, by scraping. It was shown by analysis by Messrs. Johnson, Matthey to contain 0·09% of nitrogen, with small quantities of magnesium and strontium as the major metallic impurities.

A procedure similar to that used for measurements on magnesium was adopted in the case of calcium. A new graphite viscometer was filled with calcium and set up in the vacuum enclosure, care being taken to ensure that the sphere occupied the same position within the furnace as did the one used with magnesium. Thus the previous calibration of the thermocouple F could be used again. The damping of the motion of the viscometer with the calcium in the solid state was measured for two values of the furnace current, and gave two points lying very close to the upper curve of figure 3. Since it was desirable to complete the work on calcium as quickly as possible, in view of the reactive nature of the material, it was decided to use this curve, without further measurements, to correct for the residual damping in the case of calcium.

Records of the damping were taken with the metal in the liquid state. This was done both as the metal cooled through the freezing point, and at a number of steady temperatures above this. No anomalous increase in the viscosity near the melting point was observed, but it was noted that this temperature was 812°c, a figure which was confirmed by observing the damping behaviour as the temperature was slowly increased through the melting point. While this figure is in fair agreement with a number of published results for calcium, Hoffmann and Schulze (1935) have found the melting point of pure calcium to be 851°c. From thermodynamical considerations it is difficult to see how a depression of the freezing point of about 40°c can result, as Hoffmann and Schulze claim, from about 1% of nitride impurity. Thus the question of the melting point of the pure metal can only be regarded as still unsettled. Time did not allow further refinement of the calcium used in this work, nor improvements in the techniques used in handling it, in order to investigate whether in fact the melting point could be raised to the value given by Hoffmann and Schulze.

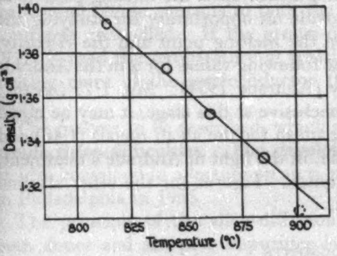


Figure 5. The density of calcium.

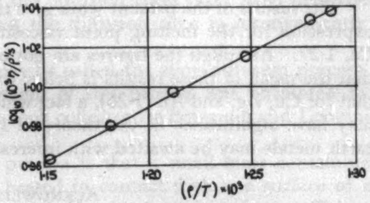


Figure 6. The experimental results for calcium fitted to Andrade's equation.

To calculate the viscosity it was necessary to measure the density of calcium at various temperatures above the melting point. This was done by a simple pyknometer method, employing a graphite density bottle suspended inside the enclosure described in § 2, and used for filling the graphite spheres. The volume of the graphite bottle could be calculated from its dimensions. When a cleaned block of calcium was melted in it (under vacuum conditions) a small drop of metal emerged from a hole in the top of the bottle. This could be swept aside by a suitable device, after which the system was allowed to cool. The density was calculated in the usual way from a knowledge of the weight of metal which filled the bottle at each of a rising sequence of temperatures. This was obtained from the weight of the empty bottle and the weight of the bottle plus metal after cooling. A correction was made to allow for the thermal expansion of the bottle. The results of this experiment are shown in figure 5. The point lying off the curve (broken circle) is believed to represent a measurement made after the rate of evaporation had become appreciable.

Values for the viscosity and density of calcium are given in table 2, whilst in figure 6 are plotted $\log_{10}(\eta/\rho^{1/3})$ against ρ/T. The latter curve gives for Andrade's constants b and c the values $2{\cdot}11 \times 10^{-3}$ and 1280; and, by extrapolation to the observed melting point of 812°c gives for the melting point viscosity the value 12·2 mP.

Table 2. The Viscosity and Density of Calcium

Temperature (°c)	812	816	833	851	867	883
Density (g cm^{-3})	1·397	1·393	1·378	1·362	1·348	1·334
Viscosity (mP)	12·22	12·09	11·53	11·01	10·57	10·12

§ 5. DISCUSSION

The value of the melting point viscosity of magnesium was found in this work to be 12·5 or 12·3 mP, according to the values of the density chosen. This is in fair agreement with the results of Gebhardt, Becker and Trägner (1955), who, since the present work was completed, have published the value of 13·2 mP. Their figure was obtained however by extrapolating to the melting point from 700°c and may well be in error on that account. No other work has been reported on the viscosity of calcium, so that the melting point viscosity of 12·2 mP (at 812°c) must be regarded as awaiting confirmation. It is perhaps worth pointing out that calcium may, if the influence of impurities upon the melting point is as great as Hoffmann and Schulze state, provide an opportunity for studying the simultaneous influence of impurities upon the melting point and the viscosity.

The results of the present work give the following values for α in the Andrade expression for the melting point viscosity (Andrade 1934): Mg, 1·18 (mean); Ca, 1·27. Although the figures are not conclusive at this stage, it may be noted that the mean of these results is below the mean for the alkali metals (1·40) and that for Cu, Ag, and Au (1·26), a fact which, in the light of Andrade's treatment, may have significance in the theory of liquids. Figures for the other alkaline earth metals may be awaited with interest.

ACKNOWLEDGMENTS

Thanks are due to Professor R. King for his supervision of this work, to the Department of Scientific and Industrial Research for a maintenance grant, and to the Managers of the Royal Institution for laboratory facilities.

REFERENCES

ANDRADE, E. N. DA C., 1934, *Phil. Mag.*, **17**, 497, 698.
ANDRADE, E. N. DA C., and CHIONG, Y. S., 1936, *Proc. Phys. Soc.*, **48**, 247.
ANDRADE, E. N. DA C., and DOBBS, E. R., 1952, *Proc. Roy. Soc.* A, **211**, 12.
ARNDT, K., and PLOETZ, G., 1927, *Z. phys. Chem.*, **130**, 184.
CHIONG, Y. S., 1936, *Proc. Roy. Soc.* A, **157**, 264.
CULPIN, M. F., 1957, *Proc. Phys. Soc.* B, **70**, 1069.
EDWARDS, J. D., and TAYLOR, C. S., 1923, *Trans. Amer. Inst. Min. (Metall.) Engrs*, **69**, 1074.
GEBHARDT, E., BECKER, M., and TRÄGNER, E., 1955, *Z. Metallk.*, **46**, 90.
GROTHE, H., and MANGELSDORF, C., 1937, *Z. Metallk.*, **29**, 352.
HOFFMANN, F., and SCHULZE, A., 1935, *Z. Metallk.*, **27**, 155.
KUBASCHEWSKI, O., and HÖRNLE, R., 1951, *Z. Metallk.*, **42**, 129.

RESEARCH NOTES

Alloy-diffusion : A Process for making diffused-base Junction Transistors†

By J. R. A. BEALE
Mullard Research Laboratory, Cross Oak Lane, Salfords, Surrey

MS. received 6th August 1957

It has been shown that solid diffusion of an impurity into a single crystal of a semiconducting material is, in general, a readily controllable process. The region into which the impurity diffuses can be made the base region of a transistor (Lee 1956, Tanenbaum and Thomas 1956). However, the width of this base region is determined not only by this diffusion process but also by the process used to form the emitter region. If the emitter is subsequently alloyed into the diffused region (Lee) the alloying depth must be closely controlled. If the emitter is formed by double diffusion (Tanenbaum and Thomas) the processes for making ohmic contacts to both the emitter and the base regions have to be accurately controlled. If the grown-diffused technique (Lehovac and Levitas 1957) is used the temperature at which the diffusion takes place is so near the melting point of the semiconductor that the diffusion time is inconveniently short.

The process to be described, which may be called the alloy-diffusion process, avoids these difficulties. The development of this process was prompted by an idea contained in a lecture on an n–p hook collector by Benjamin and Longini in Philadelphia in 1955.

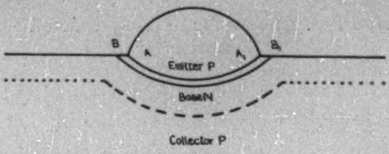
The principle of the alloy-diffusion process is that a small mass containing both donor and acceptor impurities is heated in contact with one surface of a semiconductor crystal. The temperature is maintained constant for a sufficient time for appreciable diffusion to take place from the solid–liquid interface. The impurities are chosen so that the one which is opposite in type to the original crystal diffuses much faster than the other, whereas the other is highly soluble in the semiconductor and is in sufficient concentration in the melt to ensure that on cooling the recrystallized semiconductor is dominated by this impurity.

The solid–liquid interface is therefore the reference plane, and if reproducible results are to be obtained it must be substantially stationary during the diffusion period. This can be ensured by initially heating to a higher temperature, in which case diffusion takes place from the recrystallized layer, the original solid–liquid interface being the reference plane.

The base width is therefore determined by the difference in the diffusion depths of two impurities and is practically independent of the alloying depth. In addition a reasonably substantial metal contact to the recrystallized emitter is automatically produced. Contact to the base layer is readily obtained by initially diffusing some of the fast diffusing impurity into the entire surface of the crystal, but the diffusion depth must not be much greater than the required effective base width otherwise edge effects at the emitter spoil the transistor's performance.

---

† Based on a paper read at the informal IRE/AIEE Semiconductor Device Conference at Boulder, Colorado, on July 16th 1957.



The figure illustrates an example of this process. The solid–liquid interface at the diffusion temperature is indicated by the line AA_1. The donor impurity diffuses faster and the broken line indicates the position of the collector to base junction and BB_1 indicates the emitter to base junction. The extension of the collector to base junction, indicated by the dotted lines, is due to the initial diffusion. Ohmic contact to this surface layer can be achieved by any conventional technique and the collector to base junction area can be limited by masking and etching as in the original Bell transistor (Lee 1956).

The results which will be given below were obtained by alloy-diffusing into p-type germanium a mass of some carrier metal containing about 1% Sb and 2% Ga. Lead was found to be a very suitable carrier metal. The diffusion constant of Sb is about 100 times greater than that of Ga, and Ga is more soluble in germanium than Sb, so p–n–p transistors can be obtained over a wide range of experimental conditions.

The table gives an example of what has been achieved in a group of transistors.

Parameter	Value	Remarks
α cut off frequency (f_{co}).	150–200 Mc/s.	This can be varied over a wide range by altering the manufacturing conditions.
Emitter to base depletion layer capacitance.	20–30 pF (at zero bias).	This substantially reduces the value of f_{co} when the emitter current is less than about2 mA.
Emitter to base breakdown voltage.	1·5–2 v.	
Punch through voltage.	>50 v.	
Collector capacitance.	3–4 pF (−6 v bias).	
Feedback base resistance.	15–25 Ω.	
Noise figure.	2–4 dB between 50 kc/s and 20 Mc/s.	
Neutralized power gain at 100 Mc/s.	10 dB.	
Base to collector current gain (α').	200–300.	At emitter currents of less than 100 μA, α' decreases rapidly.

The emitter and collector reverse currents were very satisfactory.

Whilst this work was proceeding we learnt that experiments along similar lines were being carried out independently by Mr. Jochems and co-workers at the Philips Research Laboratories, Eindhoven.

Acknowledgments

The author wishes to thank Mr. L. P. Morgan and Mr. L. G Cripps for making the a.c. measurements on these transistors, and Dr. K. Hoselitz for his encouragement and advice.

References

Lee, C. A., 1956, *Bell Syst. Tech. J.*, **35**, 23.
Lehovec, K., and Levitas, A., 1957, *J. Appl. Phys.*, **28**, 106.
Tanenbaum, M., and Thomas, D. E., 1956, *Bell Syst. Tech. J.*, **35**, 1.

Resistivity and Magnetoresistance of Au_2Mn

By J. H. SMITH and R. STREET
Department of Physics, The University, Sheffield

MS. received 11*th June* 1957, *and in revised form* 18*th July* 1957

Meyer and Taglang (1956) have shown, by measurements of the field dependence of magnetization and the magnetocaloric effect, that Au_2Mn possesses a property which is at present unique among metals and alloys. In zero and weak magnetic fields the material is antiferromagnetic but when the field is increased above a threshold value, which is dependent on temperature, a transition from the antiferromagnetic to a ferromagnetic state occurs. The Néel temperature, T_N, for the antiferromagnetic → paramagnetic transition and the Curie temperature, T_C, for the ferromagnetic → paramagnetic transition are both equal to 90°c. At all temperatures below 50°c the threshold field for the antiferromagnetic → ferromagnetic transition is about 15 kOe. A study of the electrical resistivity of Au_2Mn is of two-fold interest: firstly, in zero magnetic field, the temperature variation of resistivity of a metallic antiferromagnetic compound may be determined; secondly, the effect of the antiferromagnetic → ferromagnetic transition on resistivity may be found from measurements of the changes in resistance which occur when fields greater than 15 kOe are applied.

The samples of Au_2Mn used for the measurements reported here were prepared by melting together spectroscopically pure manganese and gold in an induction furnace. The ingots were then maintained at 690°c for 90 hours and quenched in water. Samples of the ingots were polished and etched in aqua regia and it was confirmed by microscopic examination that the alloy was homogeneous and single phase. Powder samples were examined by x-ray techniques and it was established that the structure was based on a body-centred tetragonal unit cell having $c/a = 2{\cdot}598_6$ and $a = 3{\cdot}368_6$ Å at 20°c. The material was highly ordered having two molecules per unit cell. X-ray photographs were also taken with the specimens maintained at 150°c and 650°c; no significant changes of structure or cell dimensions were observed. A more detailed account of this structural investigation which leads to conclusions different from those reached by Raub, Zwicker and Baur (1953) will be given elsewhere.

For resistivity measurements bar specimens of approximate dimensions $7{\cdot}0 \times 0{\cdot}8 \times 0{\cdot}8$ mm were cut from the ingots and then heated to 690°c for

90 hours and quenched in order to remove strains produced by machining. Current and potential leads were spot welded near the ends of the bars and the direction of current flow was along the length. The specimens were mounted in holders designed to fit into the tailed portion of a Dewar vessel or in an electrically heated tubular oven, both of which could be fixed when required in the gap between the poles of an electromagnet. Measurements could thus be made over the temperature range 20°K to 500°K; the maximum available magnetic fields in the gap of the electromagnet were 18 kOe with the Dewar vessel and 21 kOe with the tubular oven in position. By rotating the specimen holders the field could be applied either along or transverse to the direction of current flow. In order to correct for the effects of demagnetization, it was necessary to determine the magnetization–magnetic field curves at low temperatures to supplement the data given by Meyer and Taglang (1956). For these measurements, an ellipsoidal specimen of major axis 4 mm and minor axis 2 mm was used in conjunction with a Sucksmith ring balance adapted for use down to 20°K.

The results of the temperature variation of resistivity in zero magnetic field are plotted in figure 1: the ordinates are ρ/ρ_N and the abscissae T/T_N where ρ = resistivity at T°K and ρ_N = resistivity at the Néel temperature T_N. For this material $T_N = 366$°K and $\rho_N = 38\,\mu\Omega\,\text{cm}$. For comparison the resistivity–temperature results obtained for nickel by Potter (1937) are also plotted in normalized coordinates ρ/ρ_C and T/T_C in figure 1 (ρ_C = resistivity at the Curie temperature T_C). The similarity of the two curves suggests that the scattering of electrons from s to d states, proposed by Mott (1936), to account for the form of the temperature variation of resistivity of nickel, also occurs in antiferromagnetic Au_2Mn.

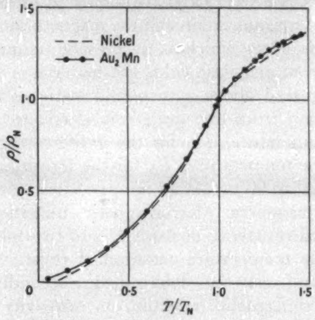


Figure 1. The temperature variation of resistivity of Au_2Mn and nickel, plotted in normalized coordinates.

In figure 2 are plotted the fractional isothermal changes in resistivity of Au_2Mn on magnetization by the magnetic field values given on the curves; the results were obtained with the field applied along the direction of current flow. The corresponding curves obtained with the field transverse to the current flow were similar in form to those in figure 2. For equal fields, corrected for demagnetization effects, the longitudinal and transverse magnetoresistance values were equal at

temperatures above T_N. Below T_N the changes in resistivity produced by transverse fields were about 15% greater in magnitude than the longitudinal values.

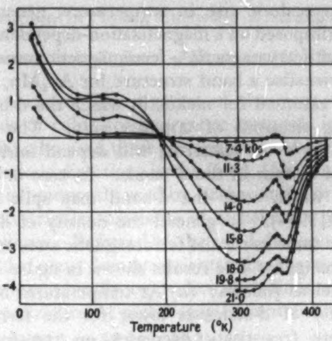
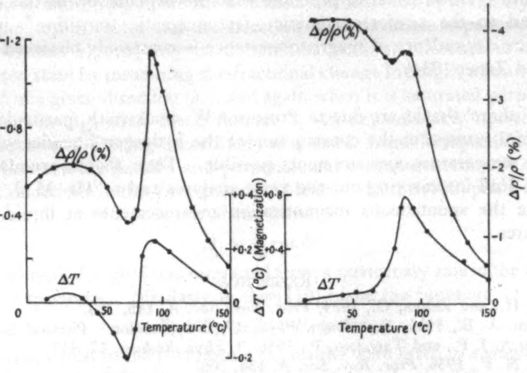


Figure 2. The temperature variation of the longitudinal magnetoresistance of Au_2Mn. The results shown correspond to magnetization by fields of the values shown.

For ferromagnetic metals and alloys, Potter (1932) showed that the resistance changes on magnetization were approximately proportional to the changes in internal energy, given by the product of the specific heat and the magnetocaloric



(*a*) (*b*)

Figure 3. Comparison of the temperature variations of magnetoresistance and the magnetocaloric temperature change for Au_2Mn. The values of $\Delta\rho/\rho$ and ΔT correspond to magnetization by fields of (*a*) 11·3 kOe and (*b*) 21 kOe.

temperature change. On magnetization of a ferromagnetic there is magnetocaloric temperature rise and an isothermal decrease in resistivity, both of which have maximum values at the Curie temperature. The close relation between the

magnetocaloric effect in Au_2Mn (Meyer and Taglang 1956) and the magnetoresistance curves given in figure 2 is apparent from figures 3 (*a*) and (*b*). Thus the magnetoresistance of Au_2Mn is considered to be a combination of an effect related to the magnetocaloric rise in temperature, generally as observed for ferromagnetics, superimposed on a magnetization-dependent change of resistance associated with the antiferromagnetic → ferromagnetic transition.

It is possible to visualize a band structure for Au_2Mn, in the ferromagnetic state, similar to that assumed for nickel, in which the d-band is split into two sub-bands containing electrons of opposite spin. The contribution to the resistivity of Au_2Mn by s → d scattering will depend *inter alia* on the density of states at the Fermi surface in the d-bands. In zero or low fields Au_2Mn is antiferromagnetic in which state the d-band may split into four sub-bands (Slater 1952, Lidiard 1954). In general the density of d-states at the Fermi surface will not be the same as for the ferromagnetic state and the resistivity will be correspondingly changed. The results shown in figure 3 may be interpreted in terms of this model as follows: (*a*) At temperatures between about 200°K and 366°K the density of d-states is lower for the ferromagnetic than the antiferromagnetic state (resistivity decreases on transformation from antiferromagnetic → ferromagnetic). (*b*) At about 200°K the densities of d-states are equal for antiferromagnetic and ferromagnetic Au_2Mn (resistivity is unchanged by antiferromagnetic → ferromagnetic transformation). (*c*) Below 200°K the density of d-states is smaller for the antiferromagnetic than for the ferromagnetic state (resistivity increases on antiferromagnetic → ferromagnetic transformation). Below 60°K, the magnetoresistance values increase rapidly with decreasing temperature. However, it is probable that the explanation of this increase is not related to the antiferromagnetic → ferromagnetic transition since similar temperature dependence of magnetoresistance is commonly observed for metals (Jones and Zener 1934).

The authors' thanks are due to Professor W. Sucksmith, particularly for his most helpful interest in the construction of the hydrogen liquefier which made the lower temperature measurements possible. Their thanks are also given to Dr. E. O. Hall for carrying out the x-ray analyses and to Mr. M. J. C. Martin who made the spontaneous magnetization measurements at liquid hydrogen temperatures.

REFERENCES

JONES, H., and ZENER, C., 1934, *Proc. Roy. Soc.* A, **145,** 268.
LIDIARD, A. B., 1954, *Rep. Progr. Phys.*, **17,** 201 (London: Physical Society).
MEYER, A. J. P., and TAGLANG, P., 1956, *J. Phys. Radium*, **17,** 457.
MOTT, N. F., 1936, *Proc. Roy. Soc.* A, **156,** 368.
POTTER, W., 1932, *Phil. Mag.*, **13,** 233; 1937, *Proc. Phys. Soc.*, **49,** 671.
RAUB, E., ZWICKER, U., and BAUR, H., 1953, *Z. Metallk.*, **44,** 312.
SLATER, J. C., 1952, *Phys. Rev.*, **82,** 538.

The Saturation Magnetostriction of Nickel Single Crystals

By R. R. BIRSS
Department of Physics, University of Nottingham

Communicated by L. F. Bates; MS. received 31st July 1957

.B
0
7

THE most comprehensive investigation of the magnetostrictive behaviour of single crystals of cubic ferromagnetic materials is that carried out by Masiyama (1928) who used three flat spheroids of nickel and measured the magnetostriction in various directions in the (100), (110) and (111) planes. The results obtained were subsequently used by Becker and Döring (1939) to evaluate the saturation magnetostriction constants of nickel by a process which may be briefly outlined as follows. Referred to the ideal demagnetized state, in which the crystal is composed of many domains orientated so that there are equal numbers in each of the crystallographically equivalent directions of easy magnetization, the saturation magnetostriction of nickel may be written in the form

$$\begin{aligned}\lambda = {} & h_1(\alpha_1^2\beta_1^2+\alpha_2^2\beta_2^2+\alpha_3^2\beta_3^2-\tfrac{1}{3})\\ & +2h_2(\alpha_1\alpha_2\beta_1\beta_2+\alpha_2\alpha_3\beta_2\beta_3+\alpha_3\alpha_1\beta_3\beta_1)\\ & +h_3(s-\tfrac{1}{3})+h_4(\alpha_1^4\beta_1^2+\alpha_2^4\beta_2^2+\alpha_3^4\beta_3^2+\tfrac{2}{3}s-\tfrac{1}{3})\\ & +2h_5(\alpha_1\alpha_2\alpha_3^2\beta_1\beta_2+\alpha_2\alpha_3\alpha_1^2\beta_2\beta_3+\alpha_3\alpha_1\alpha_2^2\beta_3\beta_1) \qquad \ldots\ldots(1)\end{aligned}$$

with

$$s=\alpha_1^2\alpha_2^2+\alpha_2^2\alpha_3^2+\alpha_3^2\alpha_1^2$$

where α_1, α_2, α_3 and β_1, β_2, β_3 are the direction cosines, relative to the crystal axes, of the magnetization vector and the direction of observation respectively.

The constants $h_1 \ldots h_5$ may be determined without reference to the initial demagnetized state by measuring the fractional change in length when the crystal is saturated in a given direction (λ_1), and again when it is saturated perpendicular to that direction (λ_2), since the difference $\lambda_1-\lambda_2$ depends only upon saturated states. Becker and Döring accordingly evaluated, from Masiyama's data, the difference between the parallel and perpendicular magnetostriction at saturation for various directions in the three principal planes. In the (100) plane the values obtained were fitted to a function of the form

$$A_1+A_2\cos 4\theta$$

where the azimuthal angle θ was measured from a quaternary axis in the equatorial plane of the specimen. Similarly in the (110) plane the function

$$B_1+B_2\cos 2\theta+B_3\cos 4\theta$$

was employed, whilst in the (111) plane the results were fitted to a single constant C_1. In this way the following constants were obtained

$$\begin{aligned}&A_1=-61\times10^{-6};\quad A_2=-14\times10^{-6};\quad B_1=-48\times10^{-6};\\ &B_2=-\ \ 5\times10^{-6};\quad B_3=-14\times10^{-6};\quad C_1=-45\times10^{-6} \qquad \ldots\ldots(2)\end{aligned}$$

Using equation (1) the quantity $\lambda_1-\lambda_2$ may be expressed as a function of θ for the three principal planes, the resulting expressions being the same as those given above with

$$\begin{aligned}A_1&=\tfrac{1}{2}h_1+\tfrac{1}{2}h_2+\tfrac{1}{2}h_4\\ A_2&=\tfrac{1}{2}h_1-\tfrac{1}{2}h_2+\tfrac{1}{2}h_4 \qquad \ldots\ldots(3\,a)\end{aligned}$$

$$
\begin{aligned}
B_1 &= \tfrac{3}{8}h_1 + \tfrac{5}{8}h_2 + \tfrac{5}{16}h_4 + \tfrac{1}{8}h_5 \\
B_2 &= \tfrac{1}{4}h_1 - \tfrac{1}{4}h_2 - \tfrac{1}{4}h_3 + \tfrac{5}{24}h_4 \\
B_3 &= \tfrac{3}{8}h_1 - \tfrac{3}{8}h_2 + \tfrac{5}{16}h_4 - \tfrac{1}{8}h_5 \\
C_1 &= \tfrac{1}{3}h_1 + \tfrac{2}{3}h_2 + \tfrac{2}{9}h_4 + \tfrac{1}{9}h_5 \qquad \ldots\ldots(3\,b)
\end{aligned}
$$

Inserting in these relations the six constants (2) derived from Masiyama's results, and assuming $h_3=0$, Becker and Döring obtained six simultaneous equations containing the four unknown quantities h_1, h_2, h_4 and h_5. They found that these equations could all be satisfied with reasonable accuracy provided that

$$
\begin{aligned}
&h_1 = -24\times10^{-6}; \quad h_2 = -47\times10^{-6}; \\
&h_4 = -51\times10^{-6}; \quad h_5 = +52\times10^{-6}. \qquad \ldots\ldots(4)
\end{aligned}
$$

These results are disturbing in several respects: it is peculiar that the fourth order constants h_4 and h_5 should both be larger than the second order constants h_1 and h_2, and none of the values obtained agrees at all well with the more recent determinations of Bozorth and Hamming (1953) who obtained

$$
\begin{aligned}
&h_1=(-68{\cdot}8\pm3{\cdot}8)\times10^{-6}; \quad h_2=(-36{\cdot}5\pm1{\cdot}9)\times10^{-6}; \\
&h_3=(-2{\cdot}8\pm3{\cdot}1)\times10^{-6}; \quad h_4=(-7{\cdot}5\pm5{\cdot}2)\times10^{-6}; \\
&h_5=(+7{\cdot}7\pm3{\cdot}1)\times10^{-6}. \qquad \ldots\ldots(5)
\end{aligned}
$$

Although Bozorth and Hamming measured only the saturation magnetostriction in a (110) plane, the technique used was much superior to that available to Masiyama and the results are correspondingly more accurate as evidenced by the magnitude of the estimated probable errors. Another unsatisfactory feature of Becker and Döring's evaluation of the magnetostriction constants is that it does not explain the periodic ($\cos 6\theta$) variation which is observed in the (111) plane, and it has been generally supposed (Bozorth 1951, Stewart 1954, Vautier 1954) that equation (1) would need to be modified to include further terms of at least sixth order in the α's.

In an attempt to improve this evaluation of the magnetostriction constants of nickel, Mason (1951) re-examined Masiyama's original data and claimed that better agreement is obtained with the values

$$
\begin{aligned}
&h_1 = -40\times10^{-6}; \quad h_2 = -46\times10^{-6}; \\
&h_4 = -36\times10^{-6}; \quad h_5 = +54\times10^{-6}. \qquad \ldots\ldots(6)
\end{aligned}
$$

Unfortunately Becker and Döring did not publish in detail their calculations for the (110) and (111) planes and the expression used by Mason to represent the variation of $\lambda_1-\lambda_2$ with azimuth in the (110) plane (eqn (20)) is incorrect. (The correct form of Mason's eqn (20) may be restored by substituting $\frac{1}{2}\sin^2\Psi$ for $\sin^2\frac{1}{2}\Psi$ throughout.) Consequently, of the four values quoted, only h_2 is necessarily correct and, in addition, the claim that higher order terms are needed in (1) to account for the magnetostriction in the (110) plane is invalidated.

Most of these anomalies surrounding the results of Masiyama are however due to mis-interpretation of the experimental data. Closer examination of the original paper reveals that the measurements were performed in the same manner as the earlier ones on iron single crystals (Honda and Masiyama 1926). Hence the quantities measured experimentally were "the longitudinal and transverse effects of magnetic expansion, that is, the components of the magnetic expansion in the direction of the field and that perpendicular to it". It is thus explicitly stated that measurements of λ were made parallel and perpendicular to the field direction and the latter measurement (λ_3) is not, in general, the same as that obtained by retaining the same measuring direction and applying a field perpendicular to it (λ_2). Moreover since λ_1 and λ_3 involve different directions of measurement the difference $\lambda_1 - \lambda_3$ is not, like $\lambda_1 - \lambda_2$, independent of the initial domain orientation.

Equation (1) may nevertheless be employed as before to calculate the variation of $\lambda_1 - \lambda_3$ with azimuth in the three principal planes. In the (100) plane $\lambda_1 - \lambda_3$ is identical in the form with $\lambda_1 - \lambda_2$, the two constants A_1 and A_2 being again given by (3 *a*) but in the (110) plane

$$\lambda_1 - \lambda_3 = B_1 + B_2 \cos 2\theta + B_3 \cos 4\theta + B_4 \cos 6\theta$$

where

$$\begin{aligned} B_1 &= \tfrac{3}{8}h_1 + \tfrac{5}{8}h_2 + \tfrac{5}{8}h_4 + \tfrac{1}{8}h_5 \\ B_2 &= \tfrac{1}{4}h_1 - \tfrac{1}{4}h_2 + \tfrac{1}{64}h_4 - \tfrac{3}{32}h_5 \\ B_3 &= \tfrac{3}{8}h_1 - \tfrac{3}{8}h_2 + \tfrac{5}{16}h_4 - \tfrac{1}{8}h_5 \\ B_4 &= \tfrac{3}{64}h_4 + \tfrac{3}{32}h_5. \end{aligned} \qquad \ldots\ldots(7)$$

In the (111) plane

$$\lambda_1 - \lambda_3 = C_1 + C_2 \cos 6\theta$$

where

$$\begin{aligned} C_1 &= \tfrac{1}{3}h_1 + \tfrac{2}{3}h_2 + \tfrac{2}{9}h_4 + \tfrac{1}{9}h_5 \\ C_2 &= -\tfrac{1}{18}h_4 - \tfrac{1}{9}h_5 \end{aligned} \qquad \ldots\ldots(8)$$

and, in this case, $\theta = 0$ in the direction of a binary axis. Thus, of the six constants (3) employed by Becker and Döring, two (B_1 and B_2) have assumed different forms and two new terms involving $\cos 6\theta$ have been added, one of which provides an explanation for the puzzling periodic variation in the (111) plane. It is possible in principle, using Becker and Döring's data (2) and the relations (3 *a*), (7) and (8), to calculate the constants $h_1 \ldots h_5$. This procedure would however be of doubtful value in view of the uncertainty associated with the demagnetized state, more especially since three separate crystals were involved in the measurements.

Nevertheless it is perhaps of interest to attempt some measure of comparison with the more recent work of Bozorth and Hamming. This may be most readily done by inserting the data of (5) in (3 *a*), (7) and (8). The resulting values are given below together with their estimated probable errors.

$$\begin{aligned} A_1 &= (-56{\cdot}4 \pm 3{\cdot}4) \times 10^{-6}; & A_2 &= (-19{\cdot}9 \pm 3{\cdot}4) \times 10^{-6}; \\ B_1 &= (-52{\cdot}3 \pm 3{\cdot}8) \times 10^{-6}; & B_2 &= (-8{\cdot}9 \pm 1{\cdot}1) \times 10^{-6}; \\ B_3 &= (-15{\cdot}4 \pm 2{\cdot}3) \times 10^{-6}; & B_4 &= (+0{\cdot}4 \pm 0{\cdot}4) \times 10^{-6}; \\ C_1 &= (-48{\cdot}1 \pm 2{\cdot}2) \times 10^{-6}; & C_2 &= (-0{\cdot}4 \pm 0{\cdot}5) \times 10^{-6}. \end{aligned}$$

It may be seen that, bearing in mind the inevitable uncertainties associated with the demagnetized states, the agreement with (2) is good and it may be concluded that equation (1) does not need to be modified by the inclusion of additional higher order terms in order to explain completely the experimental data at present available.

ACKNOWLEDGMENTS

The author would like to express his thanks to Professor L. F. Bates and Dr. E. W. Lee for their encouragement and advice.

REFERENCES

Bozorth, R. M., 1951, *Ferromagnetism* (New York : Van Nostrand).
Bozorth, R. M., and Hamming, R. W., 1953, *Phys. Rev.*, **89**, 865.
Becker, R., and Döring, W., 1939, *Ferromagnetismus* (Berlin : Springer) (Photo-lithoprint reproduction, 1943, Ann Arbor, Michigan : Edwards Brothers).
Honda, K., and Masiyama, Y., 1926, *Sci. Rep. Tôhoku Univ.*, **15**, 755.
Masiyama, Y., 1928, *Sci. Rep. Tôhoku Univ.*, **17**, 945.
Mason, W. P., 1951, *Phys. Rev.*, **82**, 715.
Stewart, K. H., 1954, *Ferromagnetic Domains* (Cambridge : University Press).
Vautier, R., 1954, *Thesis*, University of Paris.

Electric Contact to Zinc Sulphide Crystals

By G. F. ALFREY and I. COOKE
Electron Physics Department, University of Birmingham

Communicated by J. Sayers ; *MS. received* 1st *August* 1957

Smith (1955) has described a means of making ohmic contact to cadmium sulphide crystals with indium or gallium electrodes, and Kröger, Diemer, and Klasens (1956) have suggested that the electrode material diffuses into the CdS, which becomes highly n-type near the surface of the crystal, giving rise to a thin and easily penetrable barrier at the contact (ohmic contact). When small pieces of indium are placed on a zinc sulphide crystal and heated for a few seconds at about 600°c, the indium electrodes are found to make low resistance contact to the ZnS and linear voltage–current characteristics are obtained. This is in accordance with the ideas of Kröger *et al.*, but more information about conditions at the contacts can be obtained from the following experiments.

Indium electrodes were deposited on zinc sulphide by vacuum evaporation. The currents obtainable were very small and highly non-linear ($i \propto V^6$ approximately). Later experiments used thick foils of gallium and indium pressed on to the crystal instead of evaporated layers. The behaviour of foils as electrodes is identical with that of the contacts just described until high voltages are applied ($\sim$4 kv across a crystal 1 mm thick). Under these conditions the current is found to increase with time, and the crystal would be destroyed were the current not limited. After the 'breakdown' the conductivity of the specimen, measured at lower fields, is found to have increased by several orders of magnitude. Similar behaviour has been reported for cadmium sulphide (Diemer 1954, Woods 1956). The crystals studied showed electroluminescence before breakdown, but not thereafter. Without using the idea of an exhaustion barrier, it is difficult to explain electroluminescence, so it is reasonable to suppose that a contact before breakdown is associated with a barrier layer, which is modified by breakdown and afterwards plays no important part.

After breakdown the current is found to be proportional either to V or V^2, in striking contrast to the voltage–current characteristic before breakdown. Exactly similar observations were made using aluminium foil electrodes, but breakdown could not be induced with contacts of any other metals. This suggests that a group III metal is essential.

Once breakdown has occurred, using electrodes of Al, Ga, or In, it is possible to move the contact or replace the foil by any other metal, without changing the essential characteristics of the broken-down crystal. If the surface layer of the crystal next to the foil is ground away, pre-breakdown conditions are restored. These results suggest that diffusion into the crystal is the important process.

Apparently this cannot take place unless the *crystal* temperature is greater than room temperature ; if it could, no breakdown would be observable, since ohmic currents would flow from the outset. The diffusion must then require local heating, such as Joule heating in the barrier, and this could be cumulative if the current were not limited. In this way a heavily doped n-type layer could be formed at the crystal surface, as suggested by Kröger *et al.* (1956). When this condition has been realised, currents through the crystal are primarily due to electrons injected into the specimen, except at low voltages. Zinc sulphide crystals after breakdown appear to behave in this way.

Lampert (1956) has considered the problem of conduction in insulating crystals with low resistance (ohmic) contacts and has shown that the voltage at which injected carriers become important is given by

$$V_c \simeq \frac{eL^2}{2\epsilon} n \qquad \ldots\ldots(1)$$

where e is the electronic charge, n is the density of free carriers, ϵ is the permittivity of the crystal, L is the inter-electrode distance. Below this voltage, currents are characterized by the form

$$\frac{i}{A} = ne\mu\frac{V}{L} \qquad \ldots\ldots(2)$$

where μ is the electron mobility and A is the electrode area. Above V_c currents are limited by space charge and have the form

$$\frac{i}{A} = \frac{9}{8}\frac{\epsilon\mu V^2}{L^3}\,. \qquad \ldots\ldots(3)$$

Both predicted types of voltage–current characteristic have been observed in ZnS over considerable ranges of V. Observation of the transition from the linear to the quadratic form on one specimen enables numerical values to be substituted in (1) to give the free carrier concentration in zinc sulphide as about 10^{11} per cm^3.

The effective (broken-down) area of the electrode is not necessarily the same as the geometric area. An idea of the effective area can be derived from equations (2) or (3). Typical figures for one crystal range from 3×10^{-5} to the actual geometrical area of $10^{-2}\,cm^2$. The wide variation and the fact that effective areas up to the maximum value can be obtained suggest that traps in the forbidden band are without effect under the conditions of these experiments.

A method is at present being developed for applying ohmic contacts of known area to ZnS crystals. With these it will be possible to arrive at a figure for the mobility from voltage–current studies and also to carry out Hall effect measurements.

ACKNOWLEDGMENTS

We wish to thank Professor J. Sayers for placing the facilities of his department at our disposal; and the Admiralty for support of the work.

REFERENCES

DIEMER, G., 1954, *Philips Res. Rep.*, **9**, 109.
KRÖGER, F. A., DIEMER, G., and KLASENS, H. A., 1956, *Phys. Rev.*, **103**, 279.
LAMPERT, M. A., 1956, *Phys. Rev.*, **103**, 1648.
SMITH, R. W., 1955, *Phys. Rev.*, **97**, 1525.
WOODS, J., 1956, *Proc. Phys. Soc.* B, **69**, 975.

REVIEWS OF BOOKS

Tables of Functions and of Zeros of Functions. U.S. Department of Commerce, N.B.S. Applied Mathematics Series 37. Pp xi + 211. (Washington: U.S. Government Printing Office, 1954.) $2.25.

The volume contains a collection of eighteen short tables, which were published originally elsewhere. For this reason there is some diversity in the intervals of tabulation and also in the number of figures in the entries. More than half of the functions and zeros tabulated relate to Bessel functions and some of these are not easily accessible elsewhere. The functions tabulated include:

$$\int_0^x \mathrm{J}_0(t)dt; \quad \int_0^x \mathrm{Y}_0(t)dt; \quad \int_x^\infty \{\mathrm{J}_0(t)/t\}dt; \quad \tfrac{1}{2}(x)^{-n}!\,\mathrm{J}_n(x);$$

the Struve functions $L_\nu(x),\ H_\nu(x);\ x^n/n!;$

and $H(m, a, x) = \exp(-ix)F(m+l-ia,\ 2m+2;\ 2ix)$

where F is used to denote the confluent hypergeometric function ${}_1F_1$. Other tables give the first hundred Fourier coefficients of integral powers of x up to x^{10}, and sines and cosines for radian arguments between 100 and 1000. Since in this last the argument proceeds in steps of unity it is difficult to see any general use for this table.

The tables of zeros list zeros and weight factors of the first 16 Legendre polynomials and the first 15 Laguerre polynomials respectively. Apart from a table of the complex roots of $\sin z = z$, the rest of the zeros given are related to Bessel functions. These include the first five zeros of $\mathrm{J}_0(x)\mathrm{Y}_0(kx) = \mathrm{J}_0(kx)\mathrm{Y}_0(x)$, zeros of functions after fractional order and of their derivative, and the complex zeros of $\mathrm{Y}_0(z)$, $\mathrm{Y}_1(z)$ and $Y_1'(z)$.

Each table is proceeded by a concise account of the function and tabulation, which add considerable value to the tables themselves. Many of the tables included are frequently needed, and it is useful to have them published in good form in a standard series. **H. H. HOPKINS.**

Tables of the Error Function and its Derivative. U.S. Department of Commerce, N.B.S. Applied Mathematics Series 41. Pp. xi + 302. (Washington: U.S. Government Printing Office, 1954.) $3.25.

This volume is a re-issue, with a few corrections, of a table published in 1941. It represents a very comprehensive tabulation of the functions

$$H(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-\alpha^2)\,d\alpha \qquad H'(x) = \frac{2}{\sqrt{\pi}}\exp(-x^2)$$

to 15 places of decimals for the region $x = 0{\cdot}0$–56, and then to eight significant figures for $x = 4{\cdot}0$–$10{\cdot}0$. The intervals of tabulation in the different region are: 0·0000 (0·0001) 1·0000, then 1·000 (0·001) 5·600, and finally 4·00 (0·01) 10·0. Beyond $x = 5{\cdot}600$ in the 15-place part of the tables the additional entries necessary to reach $H(x) = 1$ and $H'(x) = 0$ to this accuracy are given. The layout, printing and binding are all of the high standard one now expects from the Bureau of Standards publications. **H. H. HOPKINS.**

.B
0
7

Table of the Fresnel Integral to Six Decimal Places, compiled by T. PEARCEY. Pp. 63. (Cambridge: University Press, 1956.) 12*s*. 6*d*.

In applications to theoretical physics and in particular the theory of diffraction, the Fresnel Integral is frequently needed to a precision greater than that of the four and five-figure tables which are readily available. The present tables are given to six-digit precision with seven places in the lower ranges. Tabulation is arranged so that elaborate interpolation is avoided.

The printing of the book is neat and the figures of comfortable size for easy reference.

Science and the Nation, by SIR EDWARD APPLETON. Pp. xi + 104. (Edinburgh: University Press, 1957.) 10*s*. 6*d*.

This slim little book represents the Reith Lectures for the B.B.C. almost in the form in which they were given in 1956. Sir Edward Appleton's concern is with the relations that exist between science and the nation as a whole. If, in his own words, "science is too serious a matter to be left entirely to the scientists", then someone must tell ordinary people about the way that science works, and the conditions necessary for its proper functioning. This book is one lesson in this process of educating the public.

Sir Edward is well qualified for this kind of task, since he has had experience both as an academic scientist, and at the D.S.I.R., which may be said to be the mainspring of most governmental science. He describes the threefold field of science, in university, government and industry. There does not seem to be much wrong with university science except that, as compared with industry, each experimentalist has much too little technical help from skilled technicians. Thus in universities there is an average of one technician to two or three research workers; in industry there are three or four technicians to each research worker. Yet, wisely enough, Appleton does not want to change this ratio too drastically. For to work with your hands is to learn something of your subject that you would never learn otherwise; and in universities as distinct from industry, the number and speed of obtaining results are far less important than the training of people in the methods and understanding of research.

It is in the field of industry that most change is needed. Since 97% of firms in Britain employ less than 300 workers, it is almost impossible for them to carry out much research. Yet Appleton asserts that the few big companies which do research, do so not because they are big, but they are big because they did research, even when they were small. The future of the various Research Associations is here of the greatest moment, for they can help to bring the smaller firms up to scratch. It is a grave indictment of industry that only one in five of the directors on engineering boards in Britain has any genuine scientific or technical qualification. This is likely to be the biggest single factor operative to delay the proper technological development of the original ideas. We excel in the latter: we fail miserably in the former.

Appleton has much more to say along these lines. Most of it will be familiar to those of us who are ourselves scientists. But this sort of thing needs to be said repeatedly—and Sir Edward brings into his account a number of most interesting specific illustrations. This is a useful little book, even if—from the nature of things—it is not particularly profound in its analysis, and does not break much new ground in suggested remedies. C. A. COULSON.

Friction and Lubrication, by F. P. Bowden and D. Tabor. Pp. x + 150. (London: Methuen, 1956.) 10*s.* 6*d.*

Dr. Bowden's laboratory at Cambridge has added much to our knowledge of surface phenomena, particularly in the fields of friction and lubrication. The small volume now under review, by Dr. Bowden and Dr. Tabor, is thus a survey both authoritative and concise. It will afford the practising physicist and engineer, as well as the research student, a valuable insight into the main problems of friction and lubrication; these problems are so widespread and so interesting that the book is worth reading for the general knowledge it provides. The topics range from the running of skis to the drawing of wires, from the wringing of slip gauges to the action of hydrodynamic bearings. The techniques are as simple in concept as they are fruitful in application; many of them have originated in Dr. Bowden's laboratory. Sooner or later, if the reader has anything to do with mechanisms at all, he will be glad to have read this book. R. V. JONES.

Wave Optics: Interference and Diffraction, by C. Curry. Pp. viii + 155. (London: Edward Arnold, 1957.) 21*s.*

The author's aim is to give an account of wave optics at a level intermediate between the elementary and advanced texts already available.

The field covered after a short historical introduction is roughly that of many existing treatises on Physical Optics, although consideration of polarization and birefringence is omitted entirely. This serious omission is partially compensated for by a very valuable description of optical aberrations in terms of wave front distortions.

The distinguishing feature of this book is the author's attempt to give insight into the physical principles involved by arguments of a generally non-mathematical nature. In this he succeeds in the main, but unfortunately he sometimes finds it necessary to quote without justification the results of more sophisticated treatments to fill in some of the details of his exposition. In consequence, the argument is sometimes unconvincing; an example arises in the consideration of the resolving power of a prism spectrometer. The effect on the student might be that of either a sedative or an irritant, according to disposition.

The author makes great play with the 'phase amplitude diagram', a most useful device which can give a good idea of the solution of diffraction problems where a direct analytical attack would seem forbidding. It is all the more regrettable that it should first be met with in an attempt to sum vectorially the contributions from different points in a light source.

In spite of the criticisms mentioned above, the author has produced a readable and stimulating book, which can be recommended to the first-year physics student to whom it is addressed. J. DYSON.

Analysis of Deformation, Vol. III, *Fluidity*, by K. Swainger. Pp. xxvii + 266. (London: Chapman and Hall, 1956.) 65*s.*

This third volume of the author's attempt to embrace the analysis of all continuous media in a single comprehensive theory purports to obtain the equations of hydrodynamics as entirely linear partial differential equations.

The over-simplifications of the classical theory which arise from his treatment are (i) that if **q** is the fluid velocity, then the vorticity curl **q** is constant (independent of position) and (ii) that the non-linear convective acceleration $(\mathbf{q} \cdot \nabla)\mathbf{q}$ " should not appear in the Euler acceleration term of the Navier–Stokes equations ". Since he does little more than postulate (ii) and (i) is the result of his derivation of strain-velocity from the faulty strain-displacement in his Vol. I, one can place no reliance on the theories arising from the author's devotion to the mistaken thesis that the equations of finite deformation and of hydrodynamics should be linear partial differentations. His fundamental error in Vol. I is tantamount to the erroneous supposition that an arbitrary division of a total differential increment into two parts necessarily produces total differential increments.

Whilst the book, like its predecessors, is well produced, it is nevertheless valueless. This is not a case, as the author supposes, of unorthodoxy biding its time to become a new orthodoxy, but is plain error through the author's postulation of the existence of a whole body rotational or convective displacement $\mathbf{C}^{wb}$, whereas all that really does exist is a differential element $\delta\mathbf{C}^{wb}$, becoming a total differential (and so leading to the existence of $\mathbf{C}^{wb}$) only when the displacement **D** has a constant ' vorticity ' (curl $\mathbf{D}=$const.). There is, however, nothing in the theory of a continuous medium which forces the displacement **D** or velocity **q** to possess constant vorticity and all experience goes against the truth of such an arbitrary postulate. It is reasonable to supppose that the difficulties which arise from the non-linearity of the equations of finite elastic deformations and hydrodynamics must (however slowly) be gradually overcome as we learn more of how to deal with non-linear partial differential equations of this type, and it could even be that this may entail linear processes, but there is a vast gap between this target and the author's wishful thinking hunch that these equations must be linear in character. The nature of the elementary mathematical error on which this theory is founded leave the reviewer with no grounds on which to recommend these volumes A. C. STEVENSON.

Modern Mathematics for the Engineer, edited by E. F. BECKENBACH. Pp. xx + 514. (New York, Toronto, London : McGraw-Hill, 1956.) $7.50, 56*s*. 6*d*.

This unique book is the record of an extension course on Modern Mathematics for the Engineer organised by the University of California, Los Angeles, in 1953–55, with 19 contributions by distinguished invited lecturers. The contributions are grouped under the main headings of I, Mathematical Models (7), II, Probabilistic Problems (5), III, Computational Considerations (7). It would have been evidently impossible to divide up the subject according to branches of engineering, not only because the same differential equation, for instance, may be used by the hydrodynamicist and by the electrical engineer, but also because, quietly and unobtrusively, mathematics has conquered new fields for the engineer, for which there exist as yet no professional names. In these fields which have mathematical titles such as Decision Theory, Operational Research or Dynamic Programming the engineer may still be referred to as the ' backroom boy '. ' Eminence grise ' would be a more fitting name for the boffins who are stepping on the preserves of the general and of the managing director, and who, by making a science of what used to be his art, are taking most of the fun out of it. Incidentally, this may be one of the most important and beneficial novel features of the Second Industrial Revolution.

The contributions in this book fall into three distinct classes. The first class consists of condensed reviews of more or less classical chapters of mathematics, such as theory of oscillations, calculus of variations, boundary value problems, functional transformations, conformal transformations and matrices. Richard Courant's brief and masterful review of Hyperbolic Partial Differential Equations deserved special mention, but all are highly competent.

The second class consists of articles on new mathematical branches and techniques. G. W. King's Operations Research, R. Bellmann's Dynamic Programming and G. W. Brown's Monte Carlo Methods are outstanding examples of clear, concise summaries of new branches of mathematical engineering. Four competent articles on non-linear methods, relaxation, steep descent and high-speed computers contain a wealth of practical information for the student and for the practical worker.

In the third class Norbert Wiener's article on the Theory of Prediction stands all by itself. Wiener is beyond doubt the greatest creative mathematician of the 19 authors, moreover in recent years he has whetted the appetite of spellbound audiences from London to Tokyo with tantalizing glimpses of his new general theory of linear causation. But though he has splashed his formulae on many blackboards, he has steadfastly resisted the temptation to commit them to print. Seeing at last an article from his pen the reader may well approach it with a feeling of joyful anticipation, like what he felt when as a student he opened his first book on the Calculus. But his feeling afterwards is likely to be one of bewilderment, even if his mathematical background was good enough to read the other 18 articles in the book without an effort. First, he is clubbed on the head with a mercilessly Lebesgue–Borelian introductory section. If he survives this, he is led from one abstract theorem to the other, which will probably convey no meaning to him at all, until in the very last paragraph the author suddenly descends from the clouds, and the reader shudderingly realizes that what he read was really Wiener's theory of linear causation.

Let us hope that an able expositor will be found, who will translate this mathematical gem into a language intelligible to the physicist and engineer, as Levin once translated Wiener's 'yellow peril'. As it stands it is not much more use to the engineer than if it were an anagram, or in Chinese.

Apart from the purple patch of the 20 pages of Wiener there is hardly anything in this book of 500 pages which a reader of B.Sc. standard could not digest and utilize. Applied physicists and engineers who want to keep their mathematics up to date can hardly do better than to invest in this book.

D. GABOR.

CONTENTS OF SECTION A

PAGE

Dr. P. G. **Burke** and Dr. H. H. **Robertson**. The Low Energy Elastic Scattering of Neutrons by Deuterons 777

Dr. N. **Corngold**. Slowing Down of Neutrons in Infinite Homogeneous Media 793

Dr. A. **Dalgarno** and Mr. N. **Lynn**. Properties of the Helium Atom . . . 802

Mr. T. J. M. **Boyd** and Dr. B. L. **Moiseiwitsch**. Inelastic Heavy Particle Collisions Involving the Crossing of Potential Energy Curves: V—Charge Transfer from Neutral Atoms to Doubly and Triply Charged Ions 809

Mr. W. R. S. **Garton** and Mr. A. **Rajaratnam**. Flash-Absorption Spectra of the Plasmas of Arcs and other Discharges 815

Dr. J. L. **Symonds**, Dr. J. **Warren** and Dr. J. D. **Young**. Excitation Functions up to 980 MEV for Proton-Induced Reactions in some Light Elements . . 824

Research Notes:

Dr. P. G. **Klemens**. Thermal Resistance due to Isotopic Mass Variation . 833

Mr. K. H. **Lokan**. The Integrated Cross Section from Threshold to 30 MEV for the Reaction $^9\mathrm{Be}(\gamma, 2n)\,^7\mathrm{Be}$ 836

Letters to the Editor:

Dr. N. H. **March**. The Density Matrix and the Foundations of the Statistical Theory 839

Reviews of Books 842

Contents of Section B 848

The Scattering of Radio Waves by the Moon

By J. V. EVANS
Jodrell Bank Experimental Station, University of Manchester

Communicated by A. C. B. Lovell ; MS. received 27th June 1957, and in final form 21st August 1957

Abstract. A detailed investigation has been made of the rapid fading of radio echoes from the moon on a frequency of 120 Mc/s. This has shown that the conclusion reached in an earlier paper that the moon behaves as a uniformly bright reflector is in need of revision, and that metre length radio waves are reflected from the moon, largely by a region at the centre of the visible disc, which has a radius of about one third of the moon.

This result is supported by an investigation of the shape of the received echo using a receiver bandwidth 500 c/s and transmitter pulses of 10 and 2 msec. No lengthening of the transmitted pulse could be detected, indicating that the depth of the scattering region cannot be greater than about 1 msec. Range measurements show that the echo is returned from the front edge of the moon.

The effective scattering area of a moon reflecting in this manner has been calculated, and the theoretical and observed signal to noise ratios are now in good agreement, indicating that the power reflection coefficient of the lunar surface is 0·1. The effect of this type of scattering on a moon relay communications system is briefly discussed.

§ 1. Introduction

All workers (De Witt and Stodola 1949, Kerr and Shain 1951, and Sulzer *et al.* 1952) who have obtained radio echoes from the moon have reported that they are subject to severe rapid fading. This is caused by libration which effectively rotates the moon relative to an observer, and imparts line-of-sight velocities to the scattering centres distributed over its surface. These line-of-sight velocities produce Doppler shifts in the signals reflected from the various scattering centres, causing a broadening of the power spectrum of the echoes. The shape of the power spectrum depends upon the actual distribution of the scattering centres over the disc, and thus on the particular scattering mechanism in operation.

Browne *et al.* (1956) have shown that by measuring the pulse-to-pulse autocorrelation coefficient $\gamma(f_0\tau)$ for the echoes, their radio frequency power spectrum $p(f)$ could be determined, and in this way the scattering mechanism at the surface could be investigated. The results presented in that paper were somewhat inconclusive, and this paper presents a more detailed investigation in which the autocorrelation function $\gamma(f_0\tau)$ has been determined and also the shape and range of the returned echo have been examined.

§ 2. THE DOUBLE PULSE EXPERIMENT

At a frequency of 120 Mc/s the rapid fading of moon echoes has a quasi-period of about 1 second. Thus there is little correlation between echoes about 2 seconds apart which the normal pulse recurrence frequency in use at Jodrell Bank. To derive the autocorrelation function $\gamma(f_0\tau)$ (and thence the power spectrum $p(f)$) it is necessary to use echoes spaced closer than this, at say $\frac{1}{4}$ second intervals. Unfortunately, for the present equipment, any attempt to reduce the spacing between the echoes by simply increasing their recurrence frequency would overload the transmitter. Thus the trigger circuits of the equipment were rebuilt to provide a double transmitter pulse every 4 seconds. In this way the mean power is kept constant, but the echo spacing can be varied. Transmitter pulses of 20 msec duration were used and their separation τ could be switched between $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, 1, $1\frac{1}{4}$ and $1\frac{1}{2}$ seconds. Two strobes were used to select the two echoes from the time base and these were photographed on 35 mm film, a sample of which is shown in figure 1 (Plate I).

The equipment in use at Jodrell Bank has been described previously. Each period of observation is limited by the aerial to ± 30 minutes about transit. To obtain a complete curve for $\gamma(f_0\tau)$ it is desirable to observe on all of the available pulse separations τ. However, this must be done during each period of observation, since the libration (and thus the rate of fading) changes from day to day

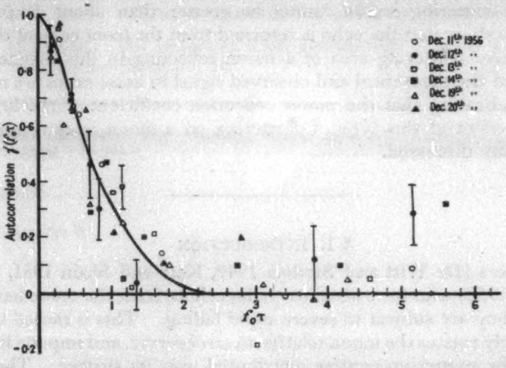


Figure 2. The values for the autocorrelation $\gamma(f_0\tau)$ plotted against $f_0\tau$ for six different nights. The full line is the mean curve through the points. Errors are shown only for December 12th, but are typical.

The observations were made during the period December 10th to December 23rd 1956. Long period fading was serious, and caused the loss of several nights and parts of others. As a result a period of about 5 minutes observation could be devoted to each of the available pulse separations every night. Thus the value for the autocorrelation $\gamma(f_0\tau)$ for any particular pulse separation τ, is the result of only about 75 pairs of echoes. Measurements were, however, only made when the echo intensity was high, so that the corrections for receiver noise are small. The results of this work are shown in figure 2, where the autocorrelation $\gamma(f_0\tau)$ is plotted against $f_0\tau$ for six different nights. The full line is the mean curve through the points. The libration l_t varied from 5 to 10×10^{-7} radian sec^{-1}

over this period, and the appropriate value of f_0 is used for each night†. If the rapid fading was caused by a mechanism other than libration the points in figure 2 would be more widely dispersed, since the rate of fading would then no longer depend upon f_0.

Most observations of moon echoes at Jodrell Bank have been made with a fixed pulse recurrence frequency of 1·8 sec. The rate of fading has been found to vary from day to day as the rate of libration l_t changes. This is shown in figure 3 where $\gamma(f_0\tau)$ for $\tau = 1{\cdot}8$ sec is plotted against l_t. Each point represents 300 pairs of echoes, and the days chosen for analysis were those on which the signal to noise ratio was high, so that the corrections for receiver noise are small.

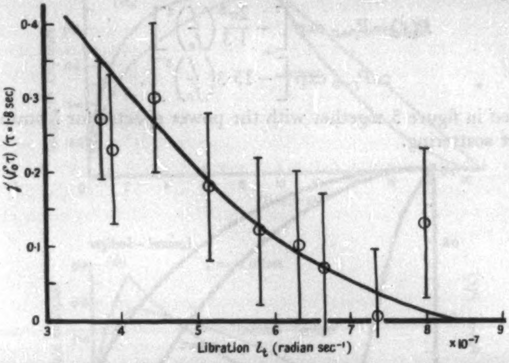


Figure 3. The variation of the autocorrelation coefficient for adjacent echoes with the total libration l_t. The full line is the mean curve of figure 2 transferred to this set of axes. Each point represents 300 pairs of echoes.

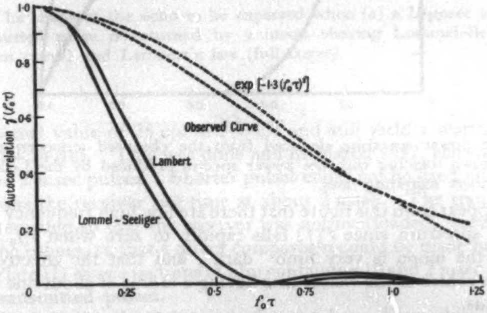


Figure 4. The observed curve for the autocorrelation function $\gamma(f_0\tau)$ shown (broken) together with the curves expected for the Lommel–Seeliger and Lambert scattering laws. Also shown is the fitted gaussian functions exp $[-1{\cdot}3(f_0\tau)^2]$.

---

† f_0 was defined in the previous paper (Browne *et al.* 1956). It is the Doppler shift produced by either limb of the moon which for a wave frequency of 120 Mc/s is equal to $l_t \times 1{\cdot}4 \times 10^8$ c/s.

The full line is the mean curve of figure 2 transferred to this set of axes. It can be seen that the results of the two experiments agree well, and that libration appears to be the mechanism responsible for the rapid fading, since the rate of fading increases with the rate of libration. In figure 4 the observed curve for $\gamma(f_0\tau)$ is shown together with the expected curves for Lommel–Seeliger and Lambert law scattering. As can be seen the observed fading is considerably slower than that which would be produced by either of these models.

For a linear detector $\gamma(f_0\tau)$ is given approximately by the square of the Fourier cosine transform of the radio frequency power spectrum. Thus since a gaussian function ($\exp[-1{\cdot}3(f_0\tau)^2]$) fits the observed results fairly closely (cf. figure 4) the power spectrum can be derived and it is found that

$$\begin{aligned} P(f) &= P_{f=0} \exp\left[-\frac{2\pi^2}{1{\cdot}3}\left(\frac{f}{f_0}\right)^2\right] \\ &\simeq P_{f=0} \exp\left[-15{\cdot}3\left(\frac{f}{f_0}\right)^2\right]. \end{aligned}$$

This is plotted in figure 5 together with the power spectra for Lommel–Seeliger and Lambert scattering.

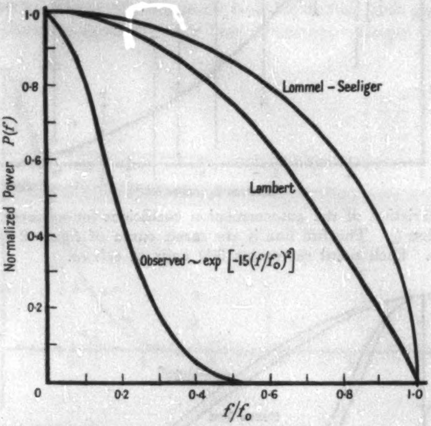


Figure 5. The power spectrum deduced from the observed autocorrelation function $\gamma(f_0\tau)$ shown together with the power spectra produced by the Lommel–Seeliger and Lambert scattering laws.

It would appear from this figure that there are few high frequency components in the power spectrum since $P(f)$ falls rapidly to zero when $f/f_0 > 0{\cdot}3$. This suggests that the moon is very limb 'dark', and that the effective scattering region is at the centre of the visible disc having a radius of about one third that of the lunar radius.

§ 3. The Pulse Shape Experiment

Kerr and Shain (1951) and also Pawsey and Bracewell (1954, p. 300) have investigated theoretically the shape of the returned echo to be expected from a moon scattering according to the Lommel–Seeliger or Lambert scattering laws. Figure 6 shows the expected echo for 10 and 2 msec transmitted pulses for these two scattering laws.

It is evident that by investigating the shape of the echo it is possible to obtain information about the scattering mechanism. Such an investigation, however, requires a receiver bandwidth of several hundred cycles at least, and this is usually much broader than the optimum and causes a serious fall in the signal to noise ratio. Kerr and Shain who tried to observe 1 msec pulses using a receiver bandwidth of 1 kc/s, state that the echoes were lengthened but they were unable to photograph any. The signal to noise ratio of the Jodrell Bank equipment is at maximum +27 dB. Thus the receiver bandwidth could be broadened from its normal value of 25 c/s to 500 c/s and still yield a worthwhile signal to noise ratio (+14 dB). This was done and in March 1957, echoes were obtained using 10 and 2 msec pulses. Shorter pulses could not be used since a bandwidth of 500 c/s sets the receiver rise time at about 2 msec. The transmitted pulses, suitably attenuated so that the receiver was not overloaded, were photographed as well as the echoes, so that a direct comparison could be made between the two. Figure 7 (Plate II) shows sections of film containing 10 and 2 msec echoes together with the transmitted pulses.

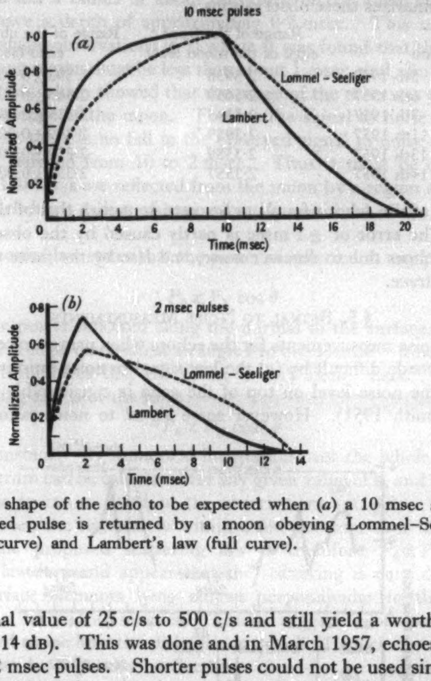


Figure 6. The shape of the echo to be expected when (*a*) a 10 msec and (*b*) a 2 msec transmitted pulse is returned by a moon obeying Lommel–Seeliger scattering (broken curve) and Lambert's law (full curve).

The echoes, unlike those drawn in figure 6 have sharp edges, and there is no detectable difference between the shape of the transmitted pulse and that of the echo. Thus it is concluded the receiver rise time is dictating the shape of both the transmitter pulse and the echo, and that the depth of the effective scattering region of the moon must be considerably less than the receiver rise time, that is of the order of 1 msec or less.

§ 4. RANGE MEASUREMENTS

Range measurements of the echoes were made, simultaneously with the pulse shape observations. A 15 kc/s crystal was used as a standard of frequency and this was divided to produce 50 c/s calibration markers on the time base. These may be seen in figure 7 (Plate II) as the negative going spikes. The 50 c/s was then further divided to provide a trigger pulse every two seconds. The transmitter pulse was delayed after this trigger pulse by about 20 msec. Since the transmitter pulse and the echo were both photographed, this delay and that produced by the receiver rise time are both removed by comparison of the two photographs. The table summarizes these observations.

Date	Range of the front edge of the moon (sec)	Range of the observed echo (sec)
March 5th 1957	2·5938	2·593 ± 0·001
March 6th 1957	2·5659	2·566 ± 0·001
March 7th 1957	2·5345	2·535 ± 0·001
March 11th 1957	2·3989	2·399 ± 0·001
March 13th 1957	2·3589	2·359 ± 0·001
March 14th 1957	2·3557	2·357 ± 0·001

The range of the echo is found to be consistent with that of the front edge of the moon. The error of ± 1 msec is partly caused by the observed spread in range of the echoes due to receiver noise, and also by the finite thickness of the photographic trace.

§ 5. SIGNAL TO NOISE MEASUREMENTS

Signal to noise measurements for the echoes when using a receiver bandwidth of 500 c/s are made difficult by (*a*) the low signal to noise ratio involved, and (*b*) the fact that the noise level on top of the echo is considerably higher than on either side (Smith 1951). However some signal to noise ratio measurements

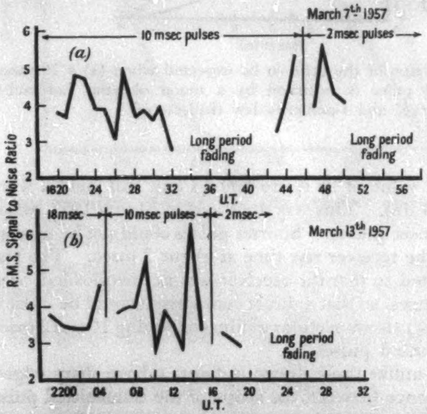


Figure 8 (*a*). The r.m.s. signal to noise ratio of strong echoes observed on March 7th 1957. The signal to noise ratio did not fall when the transmitter pulse was reduced from 10 to 2 msec. (*b*) The r.m.s. signal to noise ratio of strong echoes observed on March 13th 1957.

were made, and these are plotted in figure 8. An expected signal to noise ratio of 14 dB maximum is observed for the 10 msec pulses, but also it is found that the signal to noise ratio does not change (by more than ± 2 dB) when the transmitter pulse is reduced to 2 msec. If the moon behaved as a uniformly bright reflector a drop of 5 dB would be expected.

§ 6. DISCUSSION OF RESULTS

The investigation of the rapid fading (§ 2) shows that the moon is very limb 'dark' and that the effective scattering region is therefore at the centre of the visible disc and has a radius of about one third of the lunar radius. Such a region would have a depth of approximately 0·7 msec. This is in agreement with the pulse shape observations (§ 3) where it was found that the depth of the effective scattering region must be less than about 1 msec, and also with the range measurements (§ 4) which showed that the range of the echo was consistent with that of the front edge of the moon. Further the signal to noise measurements (§ 5) indicate that there is no fall in the observed signal to noise ratio when the pulse length is reduced from 10 to 2 msec. Thus it must be concluded that radio waves at 120 Mc/s are reflected from the moon by a region at the centre of visible disc which has a radius of approximately one third of the total radius of the moon.

Lambert's cosine law states that the power is reflected at any point according to:

$$P_\theta \propto P_N \cos\theta$$

where P_N is the power reflected along the normal to the surface, and P_θ is the power reflected along a direction at an angle θ to the normal. Since the observed power spectrum indicates that the moon is very limb 'dark', let us assume that the scattering law is of the form

$$P_\theta \propto P_N \cos^n\theta$$

where n is a constant. By numerical integration over the whole of the surface the power spectrum can be calculated for any given value of n, and then compared with the power spectrum fitted to the observations (cf. § 2). For $n=30$ the agreement between the two was to within a few per cent over the whole range $0<f/f_0<1$. The proposed scattering law is therefore $P_\theta \propto P_N \cos^{30}\theta$. By assuming this law it would appear that the scattering is very directional, and only those surface elements lying almost perpendicular to the line-of-sight reflect back in a favourable manner. This is suggestive of specular reflection though it is hard to see how the dielectric materials of which the moon's surface must consist can be 'specular' reflectors.

The results which have been presented in this paper are in disagreement with the tentative conclusion reached in the previous paper that the moon behaves as a uniformly bright reflector. Unfortunately this conclusion was based upon figures 9 and 10 of that paper which have since been found to be in error. The expected autocorrelation function $\gamma(f_0\tau)$ was shown for three types of scattering, viz. (*a*) Lommel–Seeliger (*b*) Lambert (*c*) limb brightening in figure 3, where f_0 is the Doppler shift produced by either limb and is defined in c/s. In figure 3, f_0 was given in radians and thus the ordinates of the abscissa of this figure should be divided by 2π. As a result the curves in figures 9 and 10 which show the expected autocorrelation for a uniformly bright disc are in error, since these have been derived from figure 3. Thus when the curves are drawn correctly it is found that the observed correlation is higher than expected on the basis of this model, and the conclusion reached in that paper that the moon behaves as a uniformly bright reflector is no longer tenable.

It is surprising to find that at metre length radio waves the moon is very limb 'dark' since both optically (Markov 1948) and in the infra red (Pettit and Nicholson 1930) the moon is approximately uniformly bright. Further the rapid fading measurements by Kerr and Shain (1951) who used a frequency of about 20 Mc/s gave fair agreement with a uniformly bright model. It seems unlikely that the reflection can be essentially 'specular' at 120 Mc/s, and yet not at 20 Mc/s.

§ 7. THE REFLECTION COEFFICIENT OF THE LUNAR SURFACE

Signal to noise measurements of moon echoes yield the product $g\rho$ where g is the directivity of the moon derived by Browne *et al.* (1956) and ρ its power reflection coefficient. The directivity g has been calculated by Grieg, Metzger and Waer (1948) for a moon obeying Lambert's cosine law. By replacing $\cos\theta$ in these calculations with $\cos^{30}\theta$ the directivity of the moon scattering in this manner is found to be 1·8. The theoretical estimates of the signal to noise ratio described in the previous paper were based on a value for g of 5·7, and thus are in error by about −5 dB. The discrepancy between the observed and predicted signal to noise ratios is thus reduced from -5 ± 3 dB to 0 ± 3 dB. Hence the reflection coefficient must be in the region of 0·1 as assumed in the calculations.

§ 8. CONCLUSION

Since the effective depth of the moon is 1 msec or less it becomes possible to use the moon in a communication circuit with modulation frequencies up to ~1000 c/s. This is probably just sufficient for intelligible speech and could be used for teletype. The problem of long period fading could be overcome by using circular polarization, but that of rapid fading has no easy solution. Perhaps the best approach would be to use a very high signal to noise ratio such that the amount of time during which the signal is below a certain threshold level is small. The modulation frequencies introduced by the rapid fading itself could be removed if narrow band frequency modulation were employed permitting the use of an amplitude limiter at the receiver.

ACKNOWLEDGMENTS

I am indebted to Professor A. C. B. Lovell for his encouragement and the use of the facilities at Jordell Bank, to Dr. I. C. Browne, Dr. J. G. Davies, Mr. R. Hanbury Brown and Mr. B. Rowson for their interest and advice, and to D.S.I.R. for a maintenance grant during 1954–1957.

REFERENCES

BROWNE, I. C., EVANS, J. V., HARGREAVES, J. K., and MURRAY, W. A. S., 1956, *Proc. Phys. Soc.* B, **91,** 901.

DE WITT, J. H., and STODOLA, E. K., 1949, *Proc. Inst. Radio Engrs. N.Y.*, **37,** 229.

GRIEG, D. D., METZGER, S., and WAER, R., 1948, *Proc. Inst. Radio Engrs, N.Y.*, **36,** 652.

KERR, F. J., and SHAIN, C. A., 1951, *Proc. Inst. Radio Engrs, N.Y.*, **39,** 230.

MARKOV, A. V., 1948, *Astr. J. Moscow*, **25,** 172.

PAWSEY, J. L., and BRACEWELL, R. N., 1954, *Radio Astronomy*, p. 300 (Oxford : Clarendon Press).

PETTIT, E., and NICHOLSON, S. B., 1930, *Astrophys. J.*, **71,** 102.

SMITH, R. A., 1951, *Proc. Inst. Elect. Engrs*, **98,** Part IV, 43.

SULZER, P. G., MONTGOMERY, G. F., and GERKES, I. H., 1952, *Proc. Inst. Radio Engrs*, **40,** 361.

The Thermal Expansion of Iron Ditelluride

By J. P. LLEWELLYN and T. SMITH
Marischal College, University of Aberdeen

MS. received 31*st May* 1957, *and in revised form* 28*th August* 1957

Abstract. The variation with temperature of the unit cell dimensions of two samples of orthorhombic iron ditelluride $\mathrm{FeTe}_{2\cdot 00}$ and $\mathrm{FeTe}_{2\cdot 10}$, has been measured between 17°c and 550°c using the x-ray powder method. It is found that Grüneisen's relation between the volume expansion coefficient and specific heat is adequate for $\mathrm{FeTe}_{2\cdot 00}$, but Grüneisen and Goens' theory of the linear expansion coefficients is not satisfactory. A comparison between the expansion rates of $\mathrm{FeTe}_{2\cdot 00}$ and FeSb_2 lends weight to the proposed existence of additional metalloid–metalloid bonds in the latter compound. The difference between the expansion curves of $\mathrm{FeTe}_{2\cdot 00}$ and $\mathrm{FeTe}_{2\cdot 10}$ cannot be explained in terms of additional covalent bonds and must be attributed to an increase in the ionic character of the bonds, which is associated with the vacancies present at the Te-rich composition.

§ 1. Introduction

Magnetic measurements, to be discussed elsewhere, on the marcasite-type phase (FeTe_2) of the iron–tellurium system are consistent with the presence of an antiferromagnetic Néel point at 80°K. Above this temperature the Curie–Weiss law of susceptibility is obeyed until a temperature of about 300°K is reached. Thereafter the susceptibility increases above the Curie–Weiss value, reaching a maximum at the melting point, 1050°K. In order to check whether this departure from the Curie–Weiss law could be explained in terms of a two-phase region on the phase diagram, it was decided to take a series of high temperature x-ray powder photographs of samples of iron ditelluride. From these photographs the thermal expansion has been measured. As expansion data already exist for the closely related loellingite compound, FeSb_2 (Rosenquist 1953), it was hoped that a comparison with FeTe_2 would provide a test of the possible bonding difference between the marcasite and loellingite structures. As our expansion measurements were made on a sample of stoichiometric $\mathrm{FeTe}_{2\cdot 00}$ and on a sample taken from the Te-rich end of the phase field, $\mathrm{FeTe}_{2\cdot 10}$, the effect of defects on the expansion has been obtained.

§ 2. X-Ray Measurements

2.1. *Apparatus*

A 19 cm high temperature Unicam camera was used. The temperature of the specimen was measured with the radial thermocouple, as suggested by Berry, Henry and Raynor (1951). The error in the temperature, estimated from the work of these authors, is probably as high as ± 10°c at 670°c, the highest temperature used, falling to ± 1°c at 200°c, due to uncertainty in the value for the emissivity of the specimen. It was assumed that the emissivity corresponded to that of 'black silver'.

The compounds were prepared by heating the appropriate quantities of spectrographically standardized iron and tellurium in evacuated silica tubes at

950°C for 24 hours; the temperature was then reduced to room temperature over a period of about three days. The X-ray photographs were taken of samples contained in evacuated silica capillary tubes of internal diameter 0·32 mm ($FeTe_{2\cdot 00}$) and 0·41 mm ($FeTe_{2\cdot 10}$) and wall thickness 0·06 mm. The choice of radiation was limited both by absorption in the silica and the unit cell size: zinc radiation was found to be the most satisfactory that was available. With this radiation the quality of the pictures is not good, however, because the linear absorption in the specimen itself is so high; in fact μr is of the order of 20, whereas the most suitable value is about 2. A β-filter, in the form of copper foil 0·0025 cm thick was attached round the circumference of the camera just inside the film; this had the additional advantage of acting as a screen for the fluorescent radiation. Under these conditions the exposures were about 300 mAhr, using Ilford Industrial G film.

2.2. *Estimation of Lattice Parameters*

The positions of the lines were measured on a vernier scale reading to the nearest 0·02 mm, and each angle was calculated from the mean of six readings, three on each film. The error in the Bragg angle of a line, due to errors in measurement, is less than 0·02°. Preliminary indexing of the lines was carried out by comparison with the results given by Grønvold, Haraldsen and Vihovde (1954); these authors conclude that the compound is of the marcasite type with atoms in the following positions:

2 Fe atoms at $0, 0, 0$; $\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$

4 Te atoms at $x, y, 0$; $\bar{x}, \bar{y}, 0$; $\frac{1}{2}-x, \frac{1}{2}+y, \frac{1}{2}$; $\frac{1}{2}+x, \frac{1}{2}-y, \frac{1}{2}$

where $x=0{\cdot}365$ and $y=0{\cdot}225$. The photograph taken at 670°C is of the hexagonal, NiAs type phase.

In calculating the lattice parameters, allowance was made for the displacement of the lines on the powder photographs caused by absorption in the specimen. Thus the corrected value of the Bragg angle θ_c is given by

$$\theta_c = \theta_0 - \phi \qquad \ldots\ldots(1)$$

where θ_0 is the observed Bragg angle and ϕ the correction term. Taylor and Sinclair (1945) give the expression

$$\Delta r = a(\cos\theta + \sin 2\theta/2\theta), \quad a = \text{const.} \qquad \ldots\ldots(2)$$

for the displacement Δr, caused by absorption, of a line at a Bragg angle θ. Thus ϕ is given by

$$\phi = b(\cos\theta + \sin 2\theta/2\theta), \quad b = \text{const.} \qquad \ldots\ldots(3)$$

Then, for the line $(h_i k_i l_i)$, in the case of the orthorhombic phase

$$\sin^2(\theta_0 - \phi) = A_0 h_i^2 + B_0 k_i^2 + C_0 l_i^2 \qquad \ldots\ldots(4)$$

where $A_0 = \lambda^2/4a^2$, $B_0 = \lambda^2/4b^2$, $C_0 = \lambda^2/4c^2$.

Since ϕ is less than 0·003 rad this expression can be expanded to give

$$\sin^2\theta_0 = A_0 h_i^2 + B_0 k_i^2 + C_0 l_i^2 + D_0 \frac{1}{2}\left(\frac{\cos^2\theta}{\sin\theta} + \frac{\cos^2\theta}{\theta}\right)\sin^2\theta,$$

$$D_0 = \text{const.} \qquad \ldots\ldots(5)$$

using the value of ϕ given in (3). D_0 was estimated (Taylor and Sinclair 1945) to be $0{\cdot}0023 \pm 0{\cdot}0003$ for the sample of $FeTe_{2\cdot 00}$, and $0{\cdot}0029 \pm 0{\cdot}0003$ for the sample of $FeTe_{2\cdot 10}$: these values were used in the subsequent calculations. D_0 will not change significantly as the temperature rises, as, although a rise in temperature will produce an increase in the value of μr, the difference in the value of D_0 when μr is infinite and when $\mu r = 20$, is negligible. The best values of A_0, B_0 and C_0 for the lines used were now found by using Cohen's method of least squares. The unit cell dimensions obtained from these values are shown on figure 1. By a similar method the dimensions of the high temperature hexagonal, NiAs type phase were found to be $a = 3{\cdot}821$ Å, $c = 5{\cdot}922$ Å. The error in the estimated cell dimensions is less than 1 part in 2000 in all the above cases. The major part of this error is due to the uncertainty in the value of D_0; an error of 10% in D_0 corresponds to an error of 1 part in 3000 in the cell dimensions: errors in measurement and non-uniform film shrinkage contribute to the remaining error. The error in the expansion values will be comparatively large, of the order of 15%: however, the difference between the expansion rates of the materials which are considered, is so great that this error does not invalidate their comparison.

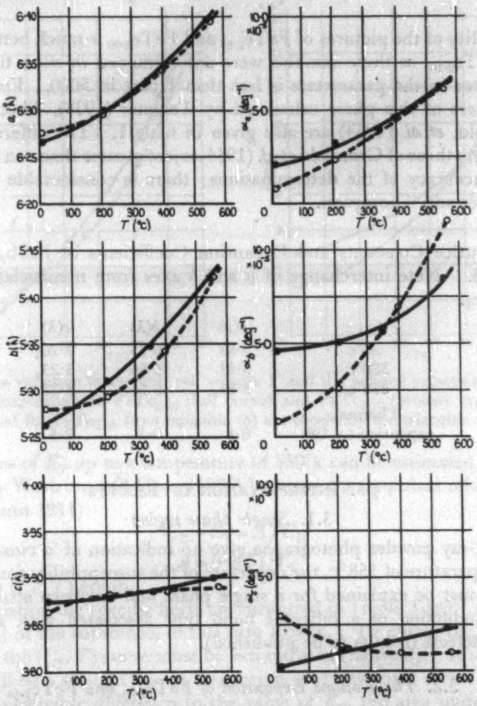


Figure 1. The variation of the lattice constants (*a*, *b*, *c*) and expansion coefficients (α_a, α_b, α_c) with temperature of $FeTe_{2{\cdot}00}$ (full curves) and $FeTe_{2{\cdot}10}$ (broken curves).

In order to find how the lattice parameters vary with composition at room temperature, powder photographs were taken of samples from the iron-rich limit of the phase field, $FeTe_{1{\cdot}95}$ and the stoichiometric composition $FeTe_{2{\cdot}00}$. The results, calculated by a method similar to that described above, but without assuming a value of D_0, are given in table 2. Also included in the table are the room temperature values to the parameters obtained for $FeTe_{2{\cdot}10}$, taken on the high temperature camera.

Table 1. Lattice Constants of Orthorhombic $FeTe_2$.

Author	Composition	*a*(Å)	*b*(Å)	*c*(Å)
Grønvold	$FeTe_{1{\cdot}90}$	$6{\cdot}264_1$	$5{\cdot}260_6$	$3{\cdot}872_4$
Present	$FeTe_{1{\cdot}95}$	6·2650	5·2639	3·8759
Grønvold	$FeTe_{2{\cdot}00}$	$6{\cdot}265_3$	$5{\cdot}265_2$	$3{\cdot}869_5$
Tengnér	,,	6·273	5·352	3·857
Thompson	,,	6·27	5·29	3·86
Present	,,	6·2655	5·2619	3·8743
Grønvold	$FeTe_{2{\cdot}10}$	$6{\cdot}274_7$	$5{\cdot}278_8$	$3{\cdot}863_6$
Present	,,	6·276	5·280	3·864

The quality of the pictures of $FeTe_{1{\cdot}95}$ and $FeTe_{2{\cdot}00}$ is much better than those taken of $FeTe_{2{\cdot}10}$, as these samples were not contained in silica tubes, and the resulting error in the parameters is less than 1 part in 5000. For comparison the parameters of this phase calculated by Tengnér (1938), Thompson (1946) and Grønvold, *et al.* (1954) are also given in table 1. The difference between our results and those of Grønvold *et al.* (1954) is just greater than can be accounted for by the accuracy of the determinations; there is considerable disagreement with Tengnér and Thompson.

Table 2. Lattice Constants and Expansion Coefficients of $FeSb_2$ (Rosenquist 1953). Note interchange of *a* and *b* axes from nomenclature used for $FeTe_2$.

	a(Å)	*b*(Å)	*c*(Å)
20°c	5·83	6·55	$3{\cdot}20_4$
380°c	5·83	6·56	$3{\cdot}23_5$
630°c	$5{\cdot}83_5$	6·56	3·25
Thermal expansion ($\times 10^5$)	0·1	0·2	2·5

§ 3. INTERPRETATION OF RESULTS

3.1. *Single phase region*

As the X-ray powder photographs give no indication of a two-phase region below a temperature of 558°c, the departure of the susceptibility from the Curie–Weiss law must be explained for a single phase and has been attributed to the gradual introduction of a different bond type associated with an increasing number of defects (results to be published).

3.2. *The Volume Expansion of* $FeTe_{2{\cdot}00}$ *and* $FeTe_{2{\cdot}10}$

Figure 2 shows the variation of the unit cell volume and volume expansion coefficient with temperature for both $FeTe_{2{\cdot}00}$ and $FeTe_{2{\cdot}10}$. This data, in conjuntion with the specific heat measurements of Westrum and Chou (1955), should provide a useful test of the validity of Grüneisen's relation for compounds of this kind over this temperature range. To make the test we shall use Grüneisen's relation in the form

$$\frac{V_T - V_0}{V_0} = \frac{E_T}{Q_0 - kE_T} \qquad \ldots\ldots(6)$$

where V_0 and V_T are the volumes at the absolute zero and T°K respectively, and $E_T = \int_0^T C_V\, dT$. Q_0 is a constant and k is a constant which is equal to $(\gamma + \frac{2}{3})$ where

$$\gamma = \frac{-V\,\partial V/\partial T}{C_v\,\partial V/\partial P}. \qquad \ldots\ldots(7)$$

The magnitudes of the quantities involved in (6) are such that at ordinary temperatures the term kE_T is small compared with Q_0; the value of V_T does not, therefore, depend critically on the value of k. This test of Grüneisen's relation will, in fact, only be able to verify the adequacy of the presence of the term kE_T, and will be no critical check on the actual form of the term itself.

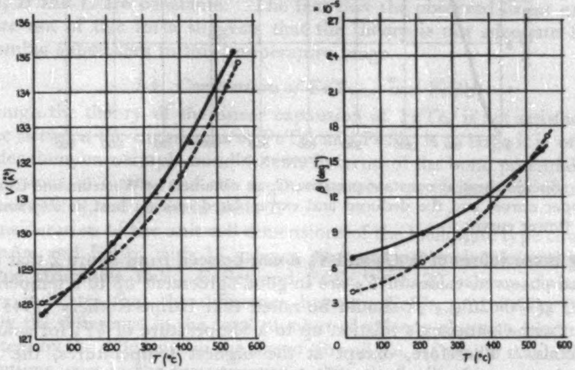


Figure 2. The variation of the unit cell volume V and the volume expansion coefficient β with temperature of $FeTe_{2\cdot00}$ (full curve) and $FeTe_{2\cdot10}$ (broken curve). Points calculated for $FeTe_{2\cdot00}$ from equation (6) are shown by the triangles.

The values of E_T up to a temperature of 350°K can be estimated from the C_P data given by Westrum and Chou (1955) by using the empirical relation (Nernst and Lindemann 1911)

$$C_P - C_V = ATC_P^2 \qquad \ldots\ldots(8)$$

$$A = 0\cdot0051/T_s \qquad \ldots\ldots(9)$$

to obtain the (C_V, T) graph.

In this relation the specific heats are measured in J mole^{-1} deg^{-1} and T_s is the melting point of the substance, in this case 1050°K. To obtain the values of E_T above 350°K the (C_V, T) curve must be extrapolated (figure 3). It is known that the graph will be a straight line in this region, and any small errors in the slope will make a negligible difference to the value of E_T, the area under the curve. The approximate value of γ, and hence k, can be found from (7) by using (9) as it is known that

$$A = \beta^2 V/XC^2 \qquad \ldots\ldots(10)$$

where β is the volume expansion coefficient, V is the molar volume and X the compressibility. By this method it is found that $k=6{\cdot}65$. The values of Q_0 and V_0 can be obtained by fitting equation (6) to the observed (V, T) graph at two points: the values obtained are $Q_0 = 1{\cdot}04 \times 10^6$ J mole^{-1} and $V_0 = 125{\cdot}92$ Å^3. This value of Q_0 agrees, within the limits of the experimental errors, with the value $1{\cdot}08 \times 10^6$ J mole^{-1} obtained from the approximate relation

$$Q_0 = C_p/\beta + 2kE_T \qquad \ldots\ldots(11)$$

where the values of C_p and β are taken at room temperature (Simon and Vohsen 1928).

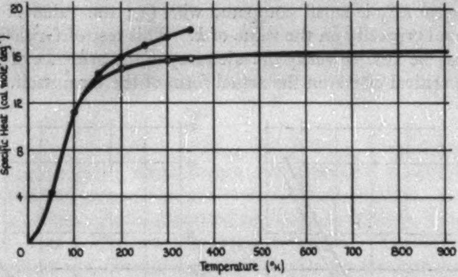


Figure 3. Specific heat at constant pressure C_p as obtained by Westrum and Chou (1955) (upper curve) and the deduced and extrapolated specific heat at constant volume C_v (lower curve).

Using these values of k, Q_0 and V_0 it can be seen from figure 2 that the predicted and observed values of V_T are in good agreement up to a temperature of about 650°K($=0{\cdot}62T_s$). It should be noted that Hume-Rothery (1945) found agreement with Grüneisen's relation up to a temperature of $\frac{2}{3}T_s$ for some hexagonal metals. Therefore, except at the highest temperatures, the relation between the expansion and specific heat can be adequately represented by equation (6). It would be interesting if the value of γ for FeTe$_{2{\cdot}10}$ could be obtained, but specific heat data is not available for this material. Although one might expect the specific heat curves for FeTe$_{2{\cdot}10}$ and FeTe$_{2{\cdot}00}$ to be closely similar it is not possible to obtain a satisfactory γ value assuming the curves to be identical, as this treatment would imply a compressibility ratio between the two materials of the order 2.

3.3. *The Variation of the Linear Expansion Coefficients with Temperatures*

For a treatment of the linear expansion coefficients α_a, α_b and α_c (see figure 1) a theory of the type derived by Grüneisen and Goens (1924) and developed by Riley (1945) will be used. It can be shown that

$$\left.\begin{aligned}
\alpha_a &= \frac{R}{V}S_{11}\gamma_x D\left(\frac{\theta_x}{T}\right) + \frac{R}{V}S_{12}\gamma_y D\left(\frac{\theta_y}{T}\right) + \frac{R}{V}S_{13}\gamma_z D\left(\frac{\theta_z}{T}\right) + LT\\
a_b &= \frac{R}{V}S_{12}\gamma_x D\left(\frac{\theta_x}{T}\right) + \frac{R}{V}S_{22}\gamma_y D\left(\frac{\theta_y}{T}\right) + \frac{R}{V}S_{23}\gamma_z D\left(\frac{\theta_z}{T}\right) + MT\\
\alpha_c &= \frac{R}{V}S_{13}\gamma_x D\left(\frac{\theta_x}{T}\right) + \frac{R}{V}S_{23}\gamma_y D\left(\frac{\theta_y}{T}\right) + \frac{R}{V}S_{33}\gamma_z D\left(\frac{\theta_z}{T}\right) + NT
\end{aligned}\right\} \quad \ldots(12)$$

where L, M and N are functions of the elastic constants S_{mn}, θ_i is a Debye temperature, γ_i is a Grüneisen number, $D(\theta_i/T)$ is the Debye function $(i=x, y, z)$, V is the gramme molecular volume and $\boldsymbol{R}$ is the gas constant. The derivation of these equations depends upon the assumptions that C_V can be split up into three terms C_{V_x}, C_{V_y} and C_{V_z}, each of which can be represented by a Debye function, and that β^2/X is constant. If one makes the reasonable assumption that the Debye temperatures θ_i are of the same order of magnitude as the Debye temperature obtained from the specific heat data, which is of the order of 200°K, then above room temperature the Debye functions will be sensibly constant. As one would expect the Grüneisen numbers also to be constant in this temperature region (Barron 1955) the expansions can be expressed in the form:—

$$\left.\begin{aligned}\alpha_a&=A+LT\\ \alpha_b&=B+MT\\ \alpha_c&=C+NT\end{aligned}\right\} \qquad \ldots\ldots(13)$$

where A, B and C are constants. The fact that the observed linear expansion curves are not of this form suggests that the theory is not adequate for these orthorhombic substances in this temperature range.

3.4. *Comparison of* $FeTe_{2\cdot00}$ *and* $FeSb_2$

Although the theory of the linear expansion of $FeTe_2$ is not satisfactory the difference between the expansions of $FeTe_2$ and $FeSb_2$ is so large it is of interest to consider them on a simple bond scheme in terms of the bond potentials. The cell dimensions of $FeSb_2$ and the mean expansion coefficients, as obtained by Rosenquist (1953), are given in table 2.

A consideration of the unit cell dimensions of the loellingite type compounds FeP_2, $FeAs_2$ and $FeSb_2$ with 18 valency electrons per molecule, and the marcasite type structures FeS_2, $FeSe_2$ and $FeTe_2$ with 20 valency electrons per molecule, indicates that the difference between $FeSb_2$ and $FeTe_2$ is caused mainly by the difference in the total number of valency electrons, and only to a lesser extent by the difference in the size of Sb and Te. If similar stable electronic configurations exist in the two structures, then the loellingite compounds must have two more bonds per molecule. If it is assumed that they are formed between neighbouring metalloid atoms, X in the *c* direction, then this lattice dimension will be reduced (Buerger 1937). If the Me–X bonds remain of constant length, this contraction in the *c*-direction will produce an expansion in the *a* and *b* directions. These changes are observed.

The interatomic distances in both $FeSb_2$ and $FeTe_{2\cdot00}$ are shown in table 3.

Table 3. Observed and Theoretical, Covalent, Interatomic distances for $FeSb_2$ and $FeTe_2$

	Me–X (Å)		X–X (Å)	
	Observed	Covalent	Observed	Covalent
$FeSb_2$	2·59	2·56	2·90	2·72
$FeTe_2$	2·57	2·60	2·93	2·74

The covalent distances included in the table were calculated using Pauling's theory of metallic bonding (Pauling 1947); the Me–X distances indicate that both $FeSb_2$ and $FeTe_{2\cdot00}$ are predominantly covalent in character.

The expansion of the compounds will now be considered in terms of this bonding scheme. As a preliminary the ratio of the expansion of isolated X–X and Me–X bonds will be found assuming a Morse potential distribution between two atoms:—

$$U(r) = D[1 - \exp\{-a(r - r_0)\}]^2 \qquad \ldots\ldots(14)$$

where D is the bond energy, r_0 is the equilibrium distance apart of the atoms and a is a constant. This equation can be expanded for small displacements $(r - r_0)$ to give

$$U(r) = Da^2(r - r_0)^2 - Da^3(r - r_0)^3$$

or

$$U(x) = fx^2 - gx^3 \qquad \ldots\ldots(15)$$

where

$$f = Da^2,\ g = Da^3,\ x = (r - r_0).$$

It is possible to make an estimate of a as follows. The force acting on an atom displaced from its equilibrium position is given by $dU(r)/dr$ and for small displacements this is approximately given by

$$dU(r)/dr \simeq 2a^2D(r - r_0) = k_0(r - r_0) \qquad \ldots\ldots(16)$$

where $k_0 = 2a^2D$ is the force constant. Badger (1935) gives an empirical formula for calculating the force constant

$$k = C_{ij}/(r - d_{ij})^3 \qquad \ldots\ldots(17)$$

where C_{ij} and d_{ij} are constants depending on the rows of the periodic table from which the atoms are taken. Thus from (16) and (17)

$$a = [C_{ij}/2(r_0 - d_{ij})^3D]^{1/2} \qquad \ldots\ldots(18)$$

The average displacement at temperature T is given by

$$\bar{x} = \int_{-\infty}^{+\infty} x \exp\{-U(x)/\boldsymbol{k}T\}\,dx \Big/ \int_{-\infty}^{+\infty} \exp\{-U(x)/\boldsymbol{k}T\}\,dx \qquad \ldots\ldots(19)$$

Using the expression given in (15) for $U(x)$, this gives

$$\bar{x} = 3\boldsymbol{k}gT/4f^2 \qquad \ldots\ldots(20)$$

Thus the coefficient of expansion is given by

$$\alpha = \frac{3\boldsymbol{k}}{4r_0}\frac{g}{f^2}.$$

Using the values of f and g given in (15) and the value of a given in (18) this reduces to

$$\alpha = \frac{3\boldsymbol{k}}{4r_0}\,2(r_0 - d_{ij})^3.$$

The ratios of the expansions of an isolated X–X bond, α_{XX}, and an Me–X bond, α_{MeX}, can therefore be found in terms of the bond energies D(XX) and D(MeX). The ratios are,

$$\left.\begin{array}{lll} (a) & \mathrm{FeSb_2} & \alpha_{\mathrm{XX}}/\alpha_{\mathrm{MeX}} = 1{\cdot}3\ \{D(\mathrm{MeX})/D(\mathrm{XX})\}^{1/2} \\ (b) & \mathrm{FeTe_2} & \alpha_{\mathrm{XX}}/\alpha_{\mathrm{MeX}} = 1{\cdot}4\ \{D(\mathrm{MeX})/D(\mathrm{XX})\}^{1/2} \end{array}\right\} \qquad \ldots\ldots(21)$$

It is possible to estimate D(MeX)/D(XX) for $\mathrm{FeSb_2}$ as the heat of formation from the elements is known to be 3·6 kcal mole^{-1} (Rosenquist 1953). From the latent heats of vaporization of Fe and Sb it can be shown that the ratio D(MeX)/D(XX) is greater than 1·5. Although the heat of formation of $\mathrm{FeTe_2}$ is not known, it would have to be greater than about 70 kcal mole^{-1} to make the ratio of the bond energies less than one. It should be noted that the Sb–Sb and Te–Te bond lengths in $\mathrm{FeSb_2}$ and $\mathrm{FeTe_2}$ are greater than the corresponding bond lengths in the elements, thus D(Sb–Sb) and D(Te–Te) as calculated from the elements are probably too large: this means that the ratio of the bond energies as estimated above is too small. It thus follows from (21) that the expansion of the X–X bonds is greater than that of the Me–X bonds.

It is interesting to note that the difference between the observed and theoretical, covalent, metalloid–metalloid distances in both loellingite and marcasite compounds decreases as the electro-negativity of the metalloid is increased (figure 4).

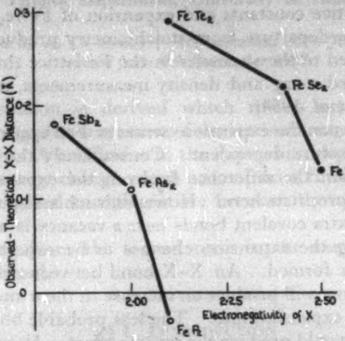


Figure 4. Discrepancy in metalloid–metalloid distances for marcasite type compounds (upper curve) and loellingite type compounds (lower curve) as a function of the electronegativity of the metalloid.

This surprising trend suggests that the large X–X distances in these compounds is due to their having a partially ionic character of the form $Me^{2-}X^{+}X^{+}$. This partial ionic character will not invalidate the conclusion that the ratio α_{XX}/α_{MeX} is greater than one, as can be shown by using a potential of the form

$$V(r) = -\frac{Ae^2z^2}{r} + \frac{Be^2}{r^n} \qquad \ldots\ldots(22)$$

in place of the Morse potential distribution given in (14).

This result will now be applied to the compounds $FeSb_2$ and $FeTe_{2\cdot00}$. In $FeTe_{2\cdot00}$ it will be assumed that Me–X bonds may be formed between an Fe atom and the 6 Te atoms that surround it octahedrally and that X–X bonds only occur between the two Te atoms in close proximity to each other. Since the X–X bonds lie in the *ab* plane an expansion of them will cause an expansion of the *a* and *b* axes; the smaller expansion of the Me–X bonds will have components along all three axes. One would expect therefore, $FeTe_{2\cdot00}$ to have large expansions along the *a* and *b* axes and a small expansion along the *c* axis.

The bonds in $FeSb_2$ are similar to those in $FeTe_{2\cdot00}$, but in addition there is an X–X bond between neighbouring metalloids in the *c* direction. If the expansion of the Me–X bonds is at first neglected, then an expansion of the X–X bonds in the *c* direction will cause a contraction of the *a* and *b* axes. This will be masked by the expansion of the X–X bond in the *ab* plane. Thus one would expect the expansion in the *a* and *b* directions to be small and that in the *c* direction to be large. As in $FeTe_{2\cdot00}$ the expansion of the Me–X bonds will produce a slight increase in the expansion along all three axes.

The expansions of $FeTe_{2\cdot00}$ and $FeSb_2$ are thus seen to be explained by the suggestion that $FeSb_2$ has two more bonds per molecule than $FeTe_2$ and these bonds are between neighbouring metalloid atoms in the *c* direction.

It may be noted that as the expansion of FeSb_2 can be accounted for in this way Rosenquist's conclusion that the large expansion along the *c* axis is an indication of antiferromagnetic coupling in this direction must be regarded as inconclusive.

3.5. *Comparison of the linear expansion coefficients of* $\mathrm{FeTe}_{2\cdot00}$ *and* $\mathrm{FeTe}_{2\cdot10}$

On figure 1 the lattice constants and expansion of $\mathrm{FeTe}_{2\cdot10}$ are also shown. It will be seen that the departure from stoichiometry produces marked effects: these must be attributed to the vacancies in the Fe lattice that are known to be present, from combined x-ray and density measurements, in concentrations of about 5% (Grønvold *et al.* 1954).

The difference between the expansion curves of $\mathrm{FeTe}_{2\cdot00}$ and $\mathrm{FeTe}_{2\cdot10}$ is seen to be markedly temperature dependent. Consequently the type of discussion that was used to explain the difference between the expansion of FeSb_2 and $\mathrm{FeTe}_{2\cdot00}$ will be less appropriate here. However, such a discussion does suggest that the formation of extra covalent bonds near a vacancy is unlikely. This can be seen by considering the expansion changes to be expected from the extra bonds which might be formed. An X–X bond between previously unbonded metalloids in the *ab* plane will produce an increase in the *a* and *b* expansion rates and a decrease in the *c* expansion rate. The less probable bond between metalloids in the *c* direction would produce the reverse effect. However the expansion curves for the two materials intersect so neither bond is sufficient explanation of the expansion changes at all temperatures. In fact magnetic measurements on these compounds (to be published) suggest an increase in the ionic character of the bonds in the neighbourhood of a vacancy. The difference in the expansion curves of $\mathrm{FeTe}_{2\cdot00}$ and $\mathrm{FeTe}_{2\cdot10}$ is to be attributed to the change in the temperature dependence of the expansion of the modified bonds.

ACKNOWLEDGMENTS

We wish to thank Professor R. V. Jones for his interest in this work, and the Department of Scientific and Industrial Research for making possible the purchase of some of the x-ray equipment.

REFERENCES

Badger, R. M., 1935, *J. Chem. Phys.*, **3**, 710.

Barron, T. H. K., 1955, *Phil. Mag.*, **46**, 720.

Berry, R. L., Henry, W. G., and Raynor, G. V., 1951, *J. Inst. Met.*, **78**, 6, 643.

Buerger, M. J., 1937, *Amer. Min.*, **22**, 48.

Grønvold, F., Haraldsen, H., and Vihovde, J., 1954, *Acta Chem. Scand.*, **8**, 1927.

Grüneisen, E., and Goens, E., 1924, *Z. Phys.*, **29**, 141.

Hume-Rothery, W., 1945, *Proc. Phys. Soc.*, **57**, 209.

Nernst, W., and Lindemann, F. A., 1911, *Z. Elektrochem.*, **17**, 817.

Pauling, L., 1947, *J. Amer. Chem. Soc.*, **69**, 542.

Riley, D. P., 1945, *Proc. Phys. Soc.*, **57**, 486.

Rosenquist, T., 1953, *Magnetic and Crystallographic Studies of the Higher Antimonides of Iron, Cobalt and Nickel* (Trondheim : Technical University of Norway).

Simon, F., and Vohsen, E., 1928, *Z. Phys. Chem.*, **133**, 165.

Taylor, A., and Sinclair, H., 1945, *Proc. Phys. Soc.*, **57**, 108, 126.

Tengnér, S., 1938, *Z. Anorg. Chem.*, **239**, 126.

Thompson, R. M., 1946, *Univ. Toronto Studies*, Geol. Ser., **51**, 35.

Westrum, E. F., and Chou, C., 1955, *Conférence de Physique de basses températures.* Paris.

Some Magnetothermal Relations for Ferromagnetics

By R. W. TEALE† and G. ROWLANDS‡
Physics Department, University of Leeds

Communicated by E. C. Stoner ; MS. received 25th July 1957, *and in revised form* 4*th September* 1957

Abstract. An expression is derived which relates associated magnetic and thermal changes for reversible rotation of the intrinsic magnetization or reversible domain wall motion. A coefficient in the expression depends upon the type of internal magnetic energy change which accompanies these processes, and hence experiments are proposed to differentiate between the various possible energy changes.

An effect arising from the second term in the magnetocrystalline energy expressions is predicted.

---

§ 1. INTRODUCTION

IN low and moderate magnetizing fields both reversible and irreversible processes contribute to changes in the magnetization of ferromagnetic specimens and to the associated changes in heat content. This is well illustrated by Bates and Sherry (1955) and Tebble and Teale (1957). The following three distinct processes contribute to these changes ; change of the magnitude of the intrinsic magnetization I_0, rotation of the direction of I_0 and movement of the walls separating ferromagnetic domains.

The results of magnetothermal measurement have been used as described below to help determine the part played by these three processes in magnetic changes.

Stoner and Rhodes (1949) derived theoretically the following expression relating the associated *reversible* magnetic and thermal changes for specimens in which these changes arise only from rotation of the intrinsic magnetization in the magnetocrystalline anisotropy and change of the magnitude of I_0.

$$\Delta Q' = ad(HI) - bHdI. \qquad \ldots\ldots(1)$$

$$a = \frac{T}{I_0}\frac{dI_0}{dT}, \quad b = \frac{T}{K_1}\frac{dK_1}{dT},$$

Q' is the heat evolved (per unit volume) in an adiabatic change, K_1 is the first magnetocrystalline anisotropy constant of the material concerned, H, I and T denote the magnetizing field (' corrected ' for the demagnetizing field due to the specimen itself), the magnetization in the field direction and the absolute temperature.

From experimental measurements, $\Delta Q'$, $d(IH)$ and HdI are deduced and a coefficient b'' is calculated where

$$b'' = \{\Delta Q' - ad(HI)\}/HdI. \qquad \ldots\ldots(2)$$

---

† Now with Mullard Research Laboratories, Salfords, Surrey.

‡ Now with the National Research Council of Canada, Ottawa.

By comparing equations (1) and (2) it can be seen that in so far as changes observed in experiment are of the type to which equation (1) applies, the coefficient b'' should be independent of H and equal to b.

Figure 1, taken from Tebble and Teale (1957), shows b'' plotted against H for an annealed specimen of nickel. The measurements were made on changes sufficiently small to be reversible. In the region above 100 Oe b'' is constant to within the accuracy of its determination and equal to b. In this region, therefore, the changes can be attributed to rotation of I_0 in the magnetocrystalline anisotropy and change of the magnitude of I_0. The difference between b'' and b in the field range 400 to 450 Oe is not significant. The probable error in b'' is large in this range.

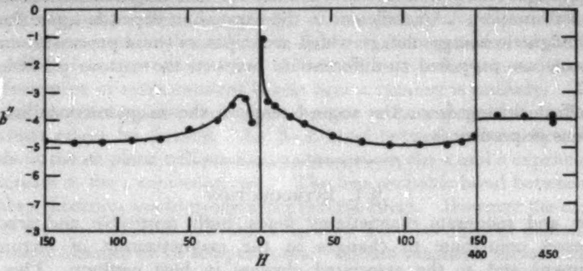


Figure 1. b'' for annealed nickel. (After Tebble and Teale 1957.) Measurements at 20·5°c in the field range ±150 Oe and at 28°c between 400 and 450 Oe. At 20·5°c $b=(T/K_1)(dK_1/dT)=-4{\cdot}8$; at 28°c $b=-5{\cdot}2$.

In lower fields however the divergence of b'' from $-4{\cdot}8$ is evident and significant. It seems likely that this arises from one or both of two sources : (*a*) The association of energetic changes other than those due to magnetocrystalline anisotropy with the rotation of I_0. (*b*) The occurrence of domain wall motion.

In the present paper relationships are derived between reversible magnetic and thermal changes for a variety of magnetic systems including some examples of rotational processes such as those mentioned in (*a*) above and of domain wall motion. The relationships allow a qualitative explanation of the low field effects mentioned above, and they suggest fresh experimental work directed towards the understanding of domain wall motion and rotational processes.

§ 2. INTERNAL FREE MAGNETIC ENERGY AND MAGNETOTHERMAL CHANGES

In this section a description is given of the relation between the magnetic part of the internal free energy of a ferromagnetic system and associated magnetothermal quantities.

When a magnetic field is applied to a ferromagnetic specimen the magnetization changes so as to produce a component in the field direction, and the field energy of the specimen, $-HI$, is thereby reduced. The equilibrium state at constant temperature under the action of a steady field H is reached when for small changes of the magnetization

$$\left(\frac{dF_M}{dI}\right)_T = H \qquad \ldots\ldots(3)$$

where F_M is the magnetic part of the internal free energy of the specimen (per unit volume). It may be associated with magnetocrystalline energy, magnetoelastic energy or several other sources which will be discussed in later sections.

Figure 2 curve *a* shows a variation on a graph in terms of which magnetic changes have often been discussed (e.g. Stoner 1950, p. 134). F_M is plotted against I, the magnetization in the field direction, at temperature T_1. Equilibrium of the system is determined by equation (3), and a small change in H at constant temperature results in the system moving along the curve so that equation (3) remains satisfied.

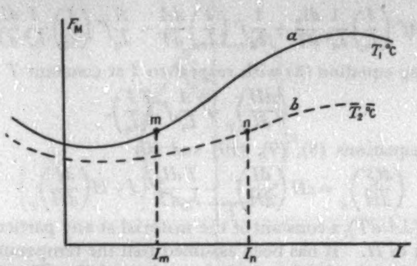


Figure 2. Magnetic part of the free internal energy of a ferromagnetic specimen plotted against I over a small range, curve *a* at T_1°c, curve *b* at T_2°c.

Suppose that whilst the system is in the state represented by the point m in figure 2 a small amount of heat is supplied to it, whilst H remains constant. The resulting change in the temperature and fundamental constants of the material (e.g. K_1, I_0) causes curve *a* to be replaced by curve *b*. The system changes reversibly to the state represented by the point n where $(dF_M/dI)_T$ is the same as at point m ; hence the magnetization changes from I_m to I_n.

The following differential relationship can easily be derived for such reversible changes :

$$\left(\frac{dI}{dT}\right)_H = -\frac{\partial^2 F_M}{\partial I \partial T}\left(\frac{dI}{dH}\right)_T. \qquad \ldots\ldots(4)$$

$(dI/dT)_H$ is connected with the heat evolved in reversible adiabatic changes by the following equation, derived thermodynamically by Stoner and Rhodes (1949).

$$\left(\frac{dQ'}{dH}\right)_S = -T\left(\frac{dI}{dT}\right)_H \qquad \ldots\ldots(5)$$

where S denotes entropy.

From equations (4) and (5)

$$\left(\frac{dQ'}{dH}\right)_S = T\frac{\partial^2 F_M}{\partial T \partial I}\left(\frac{dI}{dH}\right)_T. \qquad \ldots\ldots(6)$$

§ 3. APPLICATION TO MAGNETIC SYSTEMS

3.1. *General*

Equation (6) yields relationships between measurable quantities only when F_M has the simple form

$$F_A = Af(I/I_0) \qquad \ldots\ldots(7)$$

where A is a coefficient which is independent of I but dependent upon T, and $f(I/I_0)$ is some function of I/I_0. A general magnetothermal relation for systems to which equation (7) is applicable is derived below and a number of such systems are discussed in § 3.2.

Differentiating equation (7) with respect to I at constant T and recalling equation (3) we obtain

$$\left(\frac{dF_A}{dI}\right)_T = Af'\left(\frac{I}{I_0}\right)\frac{1}{I_0} = H. \qquad \ldots\ldots(8)$$

Differentiating equation (8) with respect to T at constant I we obtain

$$\frac{\partial^2 F_A}{\partial T \partial I} = -Af'\left(\frac{I}{I_0}\right)\frac{1}{I_0^2}\frac{dI_0}{dT} + \frac{1}{I_0}f'\left(\frac{I}{I_0}\right)\frac{dA}{dT} - \frac{A}{I_0}f''\left(\frac{I}{I_0}\right)\frac{I}{I_0^2}\frac{dI_0}{dT}. \qquad \ldots\ldots(9)$$

Differentiating equation (8) with respect to I at constant T we obtain

$$\left(\frac{dH}{dI}\right)_T = \frac{A}{I_0^2}f''\left(\frac{I}{I_0}\right). \qquad \ldots\ldots(10)$$

Combining equations (8), (9), (10) and (6)

$$\left(\frac{dQ'}{dH}\right)_S = cH\left(\frac{dI}{dH}\right)_T - \frac{T}{I_0}\frac{dI_0}{dT}\left\{I + H\left(\frac{dI}{dH}\right)_T\right\} \qquad \ldots\ldots(11)$$

where $c = (T/A)(dA/dT)$, a constant of the material at any particular temperature, and independent of H. It has been assumed that the temperature is sufficiently far below the Curie point for I_0 to be independent of H. The derivation is valid only for reversible changes.

Equation (11) is the differential form of equation (1) where b (equation (1)) has been replaced by c. The origin of the two terms on the right-hand side has been fully discussed by Stoner and Rhodes (1949) in the case of equation (1) ; we will only mention that the second term arises from the change of magnitude of I_0, and the first from whatever other magnetic change occurs (e.g. rotational change).

3.2. *Specific Magnetic Systems*

The function F_M may have various sources in real specimens. The known important sources are : (i) magnetocrystalline anisotropy, (ii) magnetostriction, (iii) magnetostatic interaction, (iv) domain wall energy. Magnetic systems are discussed below for which the form of F_M given in equation (7) is appropriate to each of these types.

3.2.1. *Rotational changes.*

Figure 3 shows a uniformly magnetized particle without domain structure. Rotation of the intrinsic magnetization is accompanied by variation of the magnetostatic energy of the particle associated with the free magnetic poles on the surface. If the variations in other forms of internal free magnetic energy which accompany the rotation of I_0 are negligible the form of F_M appropriate to the particle is $F_S = I_0^2 f(\theta)$, where θ is the angle between I_0 and H. As $I/I_0 = \cos\theta$ this becomes

$$F_S = I_0^2 g(I/I_0).$$

This form of expression for F_M is valid whatever the shape of the particle and its orientation with respect to H, though the function $g(I/I_0)$ would vary with these factors. It is therefore valid also for an assembly of such particles which do not interact.

By comparing the expression for F_S with equation (7) it can be seen that the appropriate value of A for the assembly of single domain particles is I_0^2 ; hence the appropriate value for c in equation (11) is $(2T/I_0)(dI_0/dT)$.

Alloys containing ferromagnetic impurities, powder magnets, and high coercivity alloys of the dispersion hardened type are believed to be of the form postulated above, and experiment to ascertain the degree to which the magnetothermal effects behave as predicted would be of interest.

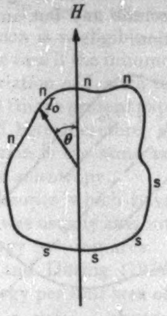


Figure 3. Single domain particle. Rotational changes are accompanied by change of magnetostatic energy.

If the dominant contribution to F_M for the single domain particles were magnetocrystalline rather than magnetostatic in origin, and if only the first term in the magnetocrystalline anisotropy expression is considerable the appropriate value for A in equation (7) becomes K_1. In equation (11) c becomes $(T/K_1)(dK_1/dT)$ and, as would be expected, the resultant magnetothermal expression is the differential form of equation (1). If magnetostrictive energy changes predominate, A becomes λ_s, the saturation magnetostrictive coefficient.

Rotational processes in which more than one type of energy change are important will be discussed in §§ 3.2.3 and 4.

3.2.2. *Domain boundary motion.*

Although the problem has been extensively investigated, the nature of the resistance to domain wall motion in ferromagnetics is still uncertain. It is evident that the presence of inhomogeneous stress and chemical impurity is of very great importance, and there exist several theories concerning the mechanism through which these factors exert their influence. We hope that the expressions developed below may contribute to the hitherto difficult task of experimental verification.

According to a very plausible theory proposed by Néel (1947, 1949) inhomogenous stress and chemical impurity cause non-zero divergence of the intrinsic magnetization within ferromagnetic domains. Resistance to the motion of domain walls then results from the change in magnetostatic energy which accompanies this motion. Lilley (1957, to be published, see also Lilley (1952)) has re-examined the treatment given by Néel and studied in detail the case where the inhomogeneities arise from stress. He finds that when it is assumed that the internal stress in a single crystal is constant in direction but with a magnitude varying according to $C=C_1+C_2\sin\lambda x\sin\mu y\sin\eta z$ the expression for the part of the internal free magnetic energy which varies with the displacement x of a 180° domain wall is

$$F_{\mathrm{D}}=\frac{DI_0^2\lambda_s^2}{K_1^2}\frac{Q^{1/2}}{(\lambda^2+Q\mu^2+Q\eta^2)(\lambda^2+Q\mu^2)^{1/2}}f(x). \qquad \ldots\ldots(12)$$

D, C_1, C_2, λ, μ and η are constants, and $Q=1+(2\pi I_0^2/K_1)$. The stress is assumed small enough to preclude any but small angles between I_0 and the magnetocrystalline easy direction.

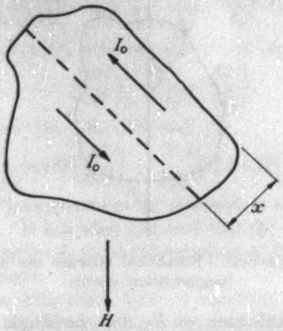


Figure 4. Particle in which domain wall motion takes place.

A single particle in which such domain wall displacement occurs is illustrated in figure 4. I/I_0 is a function of x only, and $f(x)$ in equation (12) can be replaced by $g(I/I_0)$. If it is assumed that $Q\gg 1$ equation (12) takes the form of equation (7) where $A=\alpha I_0^2\lambda_s^2/K_1^2Q$; α is a number. For iron, nickel and cobalt Q at room temperature is 40, 50 and 3 respectively. The appropriate value of c in equation (11) is

$$\frac{2T}{\lambda_s}\frac{d\lambda_s}{dT}-\frac{T}{K_1}\frac{dK_1}{dT}.$$

As in the case of the rotational model we can pass from the single particle to an assembly of particles without any change to the magnetothermal relationship.

The expression for F_{D} which results from Lilley's study of an internal stress with constant magnitude and sinusoidally varying direction is more complicated than equation (12). Nevertheless, if the assumption concerning the magnitude of Q is introduced as in the last paragraph, the expression reduces to the form of equation (7) with $A=I_0^2\lambda_s^2/K_1^2Q$ as before. We may reasonably propose that the form of c given above may be widely valid whatever the stress distribution.

As the effect of chemical impurity has been neglected in Lilley's treatment of the problem, experimental work to test the truth of the magnetothermal relation would be best carried out on a specimen in which the effect of stress would predominate. High purity nickel best fulfils this stipulation.

Néel (1947) studied the case where the local inhomogeneities in the intrinsic magnetization arise from chemical impurity or non-ferromagnetic inclusions. The magnitude of I_0 is supposed to vary according to

$$I_0 = I_m + i_\lambda \cos \eta x \sin \mu y \sin \lambda z.$$

From equations (93) and (1) in the paper by Néel the coefficient A (equation (7) of the present paper) applicable to particles in which domain wall motion is impeded by the proposed mechanism can be deduced. The value of A is $Bi_\lambda^2 K_1/I_0^2$ where B is a number. To derive this value it has been assumed that the effect of stress in the specimen is negligible.

If we suppose (as would be the case if the inhomogeneity in I_0 arose from non-ferromagnetic inclusions) the variation of i_λ with temperature is proportional to dI_0/dT, the value of c (equation (11) in the present paper) becomes $(T/K_1)(dK_1/dT)$. To the approximation employed here therefore, the magnetothermal equation for this type of domain wall motion is the same as that for rotational magnetic changes in the magnetocrystalline anisotropy.

Before Néel proposed the theories which have been discussed above, the resistance to domain wall motion was usually attributed to the effect of mechanical stress upon the domain wall energy per unit area.

This theory, which Becker and Döring (1939) have discussed at length, leads to an expression for the energy per unit area of a domain wall of the form

$$F_\lambda = 2aB(\tfrac{2}{3}\lambda_s\sigma + bK)^{1/2} \qquad \ldots\ldots(13)$$

where B is a coefficient involving the exchange energy, b is a numerical factor of the order unity, σ is the internal stress in the material where the domain wall lies, and a is the interatomic distance. B can be expressed as $N_W I_0^2$ where N_W is the Weiss molecular field constant. For specimens in which $\frac{3}{2}\lambda_s\sigma \gg bK$ equation (13) assumes the form of equation (7) with $A = N_W^{1/2} I_0 \lambda_s^{1/2} a$ (σ is a function only of domain wall displacement and may therefore be replaced by $f(I/I_0)$).

To a reasonable approximation the variation of a and N_W with temperature can be neglected in comparison with variation of λ_s and I_0, hence the appropriate form for c in equation (11) is

$$\frac{T}{I_0}\frac{dI_0}{dT} + \frac{T}{2\lambda_s}\frac{d\lambda_s}{dT}.$$

Kersten (1943) proposed that the resistance to domain wall motion is simply a result of the variation in domain wall area which must accompany the motion of the wall past a non-magnetic inclusion. In this case, if the wall is assumed to remain planar, A in equation (7) becomes γ, the domain wall energy per unit area, and c in equation (11) becomes $(T/\gamma)/(d\gamma/dT)$. This value of c is valid for any system where the relevant contribution to F_M arises from domain wall energy.

A number of workers have studied magnetic systems in which resistance to domain wall motion results from change in magnetostatic energy (e.g. Dijkstra and Wert 1950, Rhodes and Rowlands 1954, Tebble 1955). In this case A becomes I_0^2, and, as in the case of rotational changes when magnetostatic energy was predominant, c becomes $(2T/I_0)(dI_0/dT)$.

Because the exact form of the function $f(I/I_0)$ in equation (7) does not effect the magnetothermal relationship, it has been possible throughout to pass from single particle models (as in figures 3 and 4) to assemblies of such particles without change of the magnetothermal relationship. In the testing of theories of the impedance to domain wall motion this gives a great advantage to the magnetothermal method compared with the prediction and measurement of coercive fields. Such prediction often involves uncertain averaging processes. The absence from the magnetothermal relations of parameters describing the internal stress distribution and concentration of non-magnetic inclusions also greatly facilitates experimental test.

In later sections it will be seen that if domain wall motion and rotational changes occur together the effects due to the two cannot be separated. Hence the ideal specimen to use in the test would be of the ‘ picture frame ’ type employed by Williams and Shockley (1949) in which no rotational changes occur.

3.2.3. *More complicated systems.*

The magnetic systems which have been dealt with so far in this paper possessed in common two simplifying factors : (*a*) The relative magnetization I/I_0 could be expressed in terms of a single variable, θ in figure 3 and x in figure 4. (*b*) The variation of only one form of magnetic energy was taken into account. The method employed in this paper is not useful for the study of systems in which condition (*a*) does not apertain. This section concerns two systems in which condition (*a*) but not condition (*b*) holds good.

If, in a particle such as those illustrated in figures 3 and 4, the change of magnetization is accompanied by change of two forms of magnetic energy, we may express F_M in the form

$$F_M = F_1 + F_2 = Af\left(\frac{I}{I_0}\right) + Bg\left(\frac{I}{I_0}\right). \qquad \ldots\ldots(14)$$

The two terms arise from the two contributions to the magnetic energy. A and B are constants of the material and $f(I/I_0)$ and $g(I/I_0)$ are different functions of I/I_0.

If equation (14) is dealt with in a similar way to that employed in § 3.1 the result is

$$\left(\frac{dQ'}{dH}\right)_S = \frac{T}{A}\frac{dA}{dT}\left(\frac{dF_1}{dI}\right)_T\left(\frac{dI}{dH}\right)_T + \frac{T}{B}\frac{dB}{dT}\left(\frac{dF_2}{dI}\right)_T\left(\frac{dI}{dH}\right)_T - \frac{T}{I_0}\frac{dI_0}{dT}\left\{I + H\left(\frac{dI}{dH}\right)_T\right\}. \qquad \ldots\ldots(15)$$

The quantities $(dF_1/dI)_T$ and $(dF_2/dI)_T$ are not in general experimentally measurable, but we shall return to discuss equation (15) in § 4.

Another magnetic system for which the magnetothermal relationship can easily be derived is an assembly which is made up of two different types of simple particle. In one type domain wall motion occurs, and the magnetothermal relationship is equation (11) with c equal to g ; in the other type rotation of the magnetization occurs, and the magnetothermal relationship in equation (11) with c equal to e.

To determine the heat evolved in the whole system we add the amounts produced in the two types of particle. By this means we find that the magnetothermal relation for the total system is

$$\left(\frac{dQ'}{dH}\right) = gH\left(\frac{dI_B}{dH}\right)_T + eH\left(\frac{dI_R}{dH}\right)_T - \frac{T}{I_0}\frac{dI_0}{dT}\left\{I + H\left(\frac{dI}{dH}\right)_T\right\}. \qquad \ldots\ldots(16)$$

$(dI_B/dH)_T$ and $(dI_R/dH)_T$ are the contributions to the reversible susceptibility from domain wall motion and rotational processes respectively ; they add to give $(dI/dH)_T$.

In the magnetic system to which equation (16) applies, the rotational changes and domain wall motion do not interact (if they did equation (7) and therefore equation (11) would not be applicable to the two types of particle taken separately). For this reason the model and therefore equation (16) may well not give a close approximation to the behaviour of real ferromagnets.

§ 4. DISCUSSION OF EXPERIMENT

4.1. *General*

Extensive measurements on magnetothermal changes have been made by Bates and his collaborators (see for example Bates and Marshal 1953.) The method used measures the total heat changes that occur when the magnetization is changed by quite large steps. As a result, in the range of magnetizing field where irreversible changes contribute, the thermal changes measured are the sum of those due to reversible and irreversible magnetic changes. For this reason the magnetothermal relations derived above cannot be used to interpret these heat changes except in high magnetizing fields where there is no contribution from irreversible changes. In the high field range equation (1) has been used by Stoner and Rhodes and by Bates and his collaborators to interpret their results.

However, the magnetic changes studied in § 3 occur largely in the ranges of magnetizing field in which irreversible processes contribute to any large change in magnetization, and the relationships which have been derived can be of use only if the magnetothermal change associated with the reversible processes can be separated from that due to irreversible processes.

The method devised by Stoner and Rhodes and used by Tebble and Teale (1957) effects this separation in all ranges of magnetizing field, and the foregoing theory can best be compared with experiment by this method. Measurements are made of $(dI/dH)_T$ on reversible magnetic changes and ' corrected ' for the effect of the demagnetizing field as described by Tebble and Corner (1950). Measurements of $(dI/dT)_H$ using temperature changes small enough to make the changes reversible give $(dQ'/dH)_S$ through equation (5). Measured values of $(dI/dT)_H$ also need correction for the effect of the demagnetizing field, as described in Tebble, Wood and Florentin (1952). These quantities are combined to give b'' which, in the differential form of equation (2) is

$$b''=\frac{\left(\dfrac{dQ'}{dH}\right)_S-\dfrac{T}{I_0}\dfrac{dI_0}{dT}\left\{I+H\left(\dfrac{dI}{dH}\right)_T\right\}}{H(dI/dH)_T}. \qquad \ldots\ldots(17)$$

If the changes observed in experiment satisfy equation (11), b'' is independent of H and equal to c (equation (11)). In the introduction it was mentioned that magnetic changes observed in annealed specimens in moderate magnetizing fields gave $b''=(T/K_1)(dK_1/dT)$. They have been interpreted by this method in terms of the rotation of I_0 in the magnetocrystalline anisotropy.

The type of domain wall motion mentioned in §3.2.2 for which $c=(T/K_1)(dK_1/dT)$ does not provide an alternative interpretation for these moderate field changes, for in such fields domain wall motion cannot predominate.

There are no results available of measurements on specimens in which it would be expected that reversible domain wall motion would dominate the magnetothermal effects. Such measurements have been suggested in the section concerned.

Bates and Simpson (1955) made measurements of the heat evolved in adiabatic magnetic changes in materials in which it is believed that the magnetization changes by rotation in single domain particles.

However, no conclusions about the nature of the magnetic processes which occur in the low field region can be reached from the values of b'' calculated. This is because, as mentioned by Bates and Simpson, the values of b'' are subject to considerable uncertainty, and because of the contribution from irreversible processes to the heat changes measured by the method used.

Numerical values of the various coefficients referred to in the text are given in the table for iron, nickel and cobalt at 293°K. The values for $(T/\lambda_s)(d\lambda_s/dT)$ are uncertain estimates.

Metal	$\frac{T}{I_0}\frac{dI_0}{dT}$	$\frac{T}{K_1}\frac{dK_1}{dT}$	$\frac{T}{\lambda_s}\frac{d\lambda_s}{dT}$	$\frac{T}{\gamma}\frac{d\gamma}{dT}$	$\frac{2T}{\lambda_s}\frac{d\lambda}{dT}-\frac{T}{K_1}\frac{dK_1}{dT}$	$\frac{T}{I_0}\frac{dI_0}{dT}+\frac{T}{2\lambda_s}\frac{d\lambda_s}{dT}$
Iron	−0·034	−0·58	−1·5	−0·32	−2·4	−0·78
Nickel	−0·148	−4·5	−0·3	−2·4	+3·9	−0·3
Cobalt	$-0{\cdot}008_9$	$-1{\cdot}2_7$	+0·2	−0·64	+1·7	+0·1

4.2. *Low Field Effects*

A qualitative explanation of the decrease in the numerical value of b'' in figure 1 in the magnetizing field region between 100 Oe and 15 Oe can be proposed in terms of equation (16). A similar explanation applies to the measurements of Tebble and Teale (1957) on cobalt between 300 and 50 Oe.

Referring to figure 1 ; in fields above 100 Oe the constant value of b'' and its independence of H indicates that

$$\left(\frac{dI_B}{dH}\right)_T=0,\quad \left(\frac{dI_R}{dH}\right)_T=\left(\frac{dI}{dH}\right)_T$$

in this region and e (equation (16)) is $(T/K_1)(dK_1/dT)$. However, as low fields are approached $(dI_B/dH)_T$ becomes an important contribution to $(dI/dH)_T$ and if g in equation (16) is less negative than e or positive examination of equation (16) shows that a decrease in the numerical value of b'' would be expected. It can be seen from the table that all the other coefficients are numerically smaller than $(T/K_1)(dK_1/dT)$ or positive for nickel and cobalt. Hence the observed effect can be explained in terms of the occurrence of domain boundary movement in low fields.

However, we do not consider that the model of a polycrystal to which equation (16) applies justifies its use in detailed analysis of experimental results. The real situation which arises when domain wall motion occurs probably violates condition (*a*) given at the beginning of § 3.2.3.

4.3. *The Effect of the Second Magnetocrystalline Energy Term*

Equation (15) allows a discussion of the effect of the inclusion of the second term in the magnetocrystalline anisotropy expressions upon expected values of b'' for rotational processes. In the treatment given by Stoner and Rhodes only the first term was taken into account.

If this second term is taken into account the expression for F_M can be written in the form of equation (14) where $A=K_1$ and $B=K_2$. K_1 and K_2 are the first and second anisotropy constants.

Using equations (15) and (17) and recalling that

$$\left(\frac{dF_M}{dI}\right)_T=\left(\frac{dF_1}{dI}\right)_T+\left(\frac{dF_2}{dI}\right)_T=H$$

we obtain

$$b''=\frac{T}{K_1}\frac{dK_1}{dT}-\frac{1}{H}\left(\frac{dF_2}{dI}\right)_T\left\{\frac{T}{K_1}\frac{dK_1}{dT}-\frac{T}{K_2}\frac{dK_2}{dT}\right\}. \qquad \ldots\ldots(18)$$

$(1/H)(dF_2/dI)_T$ is the ratio of the change in F_2 to the change in F_M in any magnetic change ; in general it varies with H. The second term on the right-hand side of equation (18) is the deviation of b'' from the value predicted by Stoner and Rhodes for rotational processes ; it is more important the greater is K_2 compared with K_1 and the more $(T/K_2)(dK_2/dT)$ differs from $(T/K_1)(dK_1/dT)$.

Study of the explicit expressions for F_1 and F_2 for single crystals with positive uniaxial anisotropy shows that $(1/H)(dF_2/dI)_T$ increases as the intrinsic magnetization leaves the ' easy direction ' of magnetization and reaches a maximum when the crystal is magnetized in the ' hard direction '. Hence the effect upon b'' will be greatest in high fields. The opposite is true for a material with negative uniaxial anisotropy, i.e. an ' easy plane '.

For a crystal with positive uniaxial magnetocrystalline anisotropy it can be shown that when the crystal is magnetized in the hard direction

$$\frac{1}{H}\left(\frac{dF_2}{dI}\right)_T=\frac{2K_2}{K_1+2K_2}.$$

For an assembly of randomly oriented crystallites of this type (a good approximation to a polycrystal of uniaxial material in this context) the value of $(1/H)(dF_2/dI)_T$ tends in high fields towards

$$\left(0{\cdot}609+0{\cdot}813\frac{K_2}{K_1}\right)\Big/\left(1{\cdot}218+0{\cdot}533\frac{K_1}{K_2}+0{\cdot}813\frac{K_2}{K_1}\right). \qquad \ldots\ldots(19)$$

Using the measurements reported by Sucksmith and Thompson (1954) it can be estimated that the expected value of b'' decreases numerically by about 22% for a single crystal of cobalt as the magnetization rotates from the easy to the hard direction. The decrease expected for a polycrystal of cobalt in high fields was estimated as 18%. For this material the effect would not become considerable until magnetizing fields of several thousand oersteds are applied.

An examination of the explicit expressions for F_1 and F_2 for nickel shows that F_2 contributes considerably to rotational changes near the ' easy direction ', crystallographically (111). Hence the effect under discussion will be present in low fields. Until reliable measurements of K_2 and dK_2/dT are available the importance of the effect cannot be predicted. It may provide an alternative explanation for the low field variation of b'' observed in nickel.

It is possible that the variation of b'' with H in high fields, which was noted by Bates and Marshall (1953) for 4% silicon–iron, is due to the effect of K_2 ; measured values of K_2 and dK_2/dT are not available. The same effect on b'' could be produced by some energetic change other than that due to magnetocrystalline anisotropy, if this energetic change became more important as the magnetizing field increased. A qualitative description of such an effect could be given in terms of equation (15) in a similar way to the description of the effect of K_2.

§ 5. CONCLUSION

It has been shown how magnetothermal measurements on specimens in which domain wall motion or rotational changes occur in isolation from one another can give information about the type of energy changes involved. This is perhaps the most vital information in understanding these processes.

Two types of magnetic change have been studied for which the coefficient b'' has been shown to vary with magnetizing field ; and a qualitative explanation has been offered for the observed variation in low magnetizing fields.

The effect upon b'' expected due to the second term in the magnetocrystalline anisotropy expression has been indicated.

ACKNOWLEDGMENTS

The authors would like to thank Professor Stoner, Dr. Tebble and Dr. Rhodes for their interest in the work reported, Dr. Hoselitz for his help in preparing the paper for publication and Dr. Berz for deriving equation (19). Each of us is indebted to the Department of Scientific and Industrial Research for a maintenance grant.

REFERENCES

BATES, L. F., and MARSHAL, G., 1953, *Rev. Mod. Phys.*, **25**, 17.
BATES, L. F., and SHERRY, N. P. R., 1955, *Proc. Phys. Soc.* B, **68**, 642.
BATES, L. F., and SIMPSON, A. W., 1955, *Proc. Phys. Soc.* B, **68**, 849.
BECKER, R., and DÖRING, W., 1939, *Ferromagnetismus* (Berlin : Springer).
DIJKSTRA, L. F., and WERT, C., 1950, *Phys. Rev.*, **79**, 979.
KERSTEN, M., 1943, *Grundlagen einer Theorie der ferromagnetischen Hysterese und Koersitivkraft* (Leipzig : Hirzel).
LILLEY, B. A., 1952, *Thesis*, University of Leeds.
NÉEL, L., 1944, *J. Phys. Radium*, **5**, 241 ; 1947, *Ann. Univ. Grenoble*, **22**, 299 ; 1949, *Physica*, **15**, 225.
RHODES, P., and ROWLANDS, G., 1954, *Proc. Leeds Phil. Lit. Soc.*, **6**, 191.
STONER, E. C., 1950, *Rep. Progr. Phys.*, **13**, 83 (London : Physical Society).
STONER, E. C., and RHODES, P., 1949, *Phil. Mag.*, **40**, 481.
SUCKSMITH, W., and THOMPSON, J. E., 1954, *Proc. Roy. Soc.*, **225**, 362.
TEBBLE, R. S., 1955, *Proc. Phys. Soc.* B, **68**, 1017.
TEBBLE, R. S., and CORNER, W. D., 1950, *Proc. Phys. Soc.* B, **63**, 1005.
TEBBLE, R. S., and TEALE, R. W., 1957, *Proc. Phys. Soc.* B, **70**, 51.
TEBBLE, R. S., WOOD, J. E., and FLORENTIN, J. J., 1952, *Proc. Phys. Soc.* B, **65**, 858.
WILLIAMS, H. J., and SHOCKLEY, W., 1949, *Phys. Rev.*, **75**, 179.

The Effect of Liquid Media on the Grey Tin ⇌ White Tin Transformation in Commercially Pure Tin

By R. W. SMITH and G. V. RAYNOR

Department of Physical Metallurgy, University of Birmingham

MS. received 22nd July, and in final form 6th September 1957

Abstract. Dilatometric experiments using xylene, ethyl alcohol, alcoholic 'pink salt' solution and distilled water show that the rates of transformation of the $\alpha \rightarrow \beta$ and $\beta \rightarrow \alpha$ reactions in tin are affected by the surrounding medium. The $\alpha \rightarrow \beta$ rate for material which has undergone one previous $\beta \rightarrow \alpha$ change is greater in water or ethyl alcohol than in xylene. All rates increase with temperature up to the highest investigated (29·86°c), and do not reach a maximum at about 25°c as reported by earlier workers. Alcoholic 'pink salt' solution produces a much higher $\beta \rightarrow \alpha$ transformation rate than other liquids in specimens which have undergone two successive transformations. The effects of xylene and ethyl alcohol during repeated transformation were studied at 30°c and 50°c; $\alpha \rightarrow \beta$ rates were highest at the higher temperature. At 30°c, transformation rates increased with repeated transformation under xylene, but decreased under ethyl alcohol. It is suggested that xylene behaves as an inert liquid, while the effects observed with other liquids are due to the presence of a hydroxyl group. The increase of $\beta \rightarrow \alpha$ transformation rate obtained with alcoholic pink salt is considered to be due to the reducing nature of the medium, and the formation of local electrolytic cells of various types. The results of certain earlier investigations are discussed on the basis of the present work.

§ 1. INTRODUCTION

When preparing tin specimens for dilatometric examination in connection with other work in progress in the authors' laboratory, it was noted that the time required for the β (white tin) $\rightarrow \alpha$ (grey tin) transformation to proceed to apparent completion was influenced by the dilatometric liquid. In the course of experiments, glass bulbs containing white tin were sealed on to capillary tubes, and in some cases small quantities of water were formed inside the apparatus; the subsequent $\beta \rightarrow \alpha$ transformation, after inoculation with grey tin, was slower if this water was not removed than in the absence of water, whether the transformation was carried out in the presence of the liquid medium chosen as the dilatometric liquid or not. Further, repeatedly transformed dilatometric specimens showed higher dilation rates near the transformation temperature if covered by xylene than if covered by alcoholic 'pink salt' solution (a solution of $(NH_4)_2SnCl_6$ in ethyl alcohol, saturated at -40°c). These observations suggested that whether or not the liquid chosen as the dilatometric indicator was a polar compound was of importance; in view of the variety of liquids employed in earlier work on the tin transformation, an examination of the effects of various polar and non-polar liquids on the course of the tin transformation has been carried out. The experiments were designed to provide information under the following headings: (*a*) the effect of liquid media on the linear velocity of the $\beta\rightarrow\alpha$ transformation, (*b*) the effect of liquid media on the dilation velocity of the $\alpha\rightleftharpoons\beta$ transformation, (*c*) the effect of liquid media on the $\alpha\rightarrow\beta$ transformation in repeatedly transformed material.

§ 2. EXPERIMENTAL METHODS AND RESULTS

(*a*) *The Linear Velocity of the* $\beta\rightarrow\alpha$ *Transformation*

Wires were prepared by extruding pure 'Vulcan' tin (99·9972% Sn) at 100°c, giving specimens 5 cm long and 2 mm in diameter. They were annealed *in vacuo* for 60 h at 175°c and then, except for one end of each specimen, coated with a thin film of transparent lacquer. Specimens were inoculated by indenting the exposed ends with a sharpened file, previously dipped in grey tin powder. They were then placed in separate glass tubes and the appropriate liquid added. Finally the tubes were sealed, placed vertically on solid carbon dioxide in a suitable enclosure, and observed at intervals. Growth of grey tin proceeded from a single centre; when the transformation involved the entire cross section of a specimen the rate of advance of the interface was measured by observing the reduction in length of the white tin zone, using a low power measuring microscope.

The liquids used were ethyl alcohol, alcoholic pink salt solution, toluene and xylene. All were of high commercial purity but were not specially dehydrated. Although there was a general tendency for toluene to produce the highest linear transformation velocity ($0{\cdot}68\ \mathrm{mm\,h^{-1}}$) and pink salt the lowest ($0{\cdot}53\ \mathrm{mm\,h^{-1}}$) the results were not sufficiently conclusive for rigorous treatment. They suggest that the effect of liquid media on the transformation is small.

(*b*) *The Dilation Velocity of the* $\alpha\rightleftharpoons\beta$ *Transformation*

Since more sensitivity was required for investigating the effects of liquid media, a dilatometric method was developed for this portion of the research. Four 50 g samples of 'Pass-S' tin (99·9970% Sn) were weighed into small glass flasks, which were washed out with ether and pumped dry. A few grains of dry grey tin were added, and the flasks, securely stoppered, were placed on solid carbon dioxide for four months. The flasks were then allowed to warm up to room temperature (15°c), evacuated, and filled with one of the following liquids: LS1 anhydrous xylene, LS2 anhydrous ethyl alcohol, LS3 alcoholic pink salt solution, LS4 distilled water. Capillary tubes were then fitted. The xylene was dehydrated by standing over sodium for two months; water-free alcohol was prepared using magnesium and iodine in the standard Grignard reaction.

The dilatometers were secured in a thermostatically controlled enclosure (±1/200°c) and the changes in the meniscus height with time at various temperatures were measured with a vernier cathetometer. The results are given in table 1, and discussed in § 3.

At 29·86°c, six minutes elapsed before the meniscus level of each dilatometer was noted; readings were then taken at two minute intervals. After three hours specimen LS4 was considerably lighter in colour than the other specimens, and the particle size was much larger ($\sim 5\ \mathrm{mm^3}$).

After the 29·86°c treatment, all specimens were heated for 30 minutes at 45°c, before placing in the thermostat at 1·48°c, to convert the bulk of any retained grey tin to the white modification, and thus to standardize the condition of each specimen for the subsequent $\beta\rightarrow\alpha$ transition.

Table 1

(1)	(2)	(3)	(4)	(5)	(6)	(7)
			LS1	LS2	LS3	LS4
1·48	18	4	0·003	0·000	0·108	0·004
19·06	19	1	−0·052	−0·093	−0·041	−0·027
25·60	2	2	−1·03	−2·70	−0·67	−1·05
29·86	1	3	*a* −49·2	−94·8	−36·6	−108·0
			b −24·6	−35·4	−25·2	−68·4

(1) Temperature (°c) ; (2) duration of test (h) ; (3) order of test ; (4)–(7) rate of change of meniscus level (cm/hr).

a and *b* respectively denote values after 6 and 19 minutes. Negative signs indicate a falling meniscus. Between tests 2 and 3 the dilatometers were maintained overnight at 15°c, during which time very little transformation took place.

(*c*) *The $\alpha \rightarrow \beta$ Transformation in Repeatedly Transformed Material*

Two specimens of Vulcan tin, of weight 55·38 g corresponding to a volume change of 2·0 cm^3 on transformation, were sealed in Pyrex dilatometers with graduated capillary columns. The scale could be read to 0·01 cm^3, corresponding to an accuracy to $\pm\frac{1}{2}\%$ in estimation of the amount of material transformed at any given time; a volume change of 1 cm^3 produced an alteration in meniscus level of 5 cm. After inoculation each dilatometer was filled with the appropriate liquid. These were: specimen D1, commercial xylene; specimen D2, anhydrous ethyl alcohol. Sufficient liquid was used for the meniscus to lie just within the graduated region of the capillary when the vessels were placed in melting ice. This level, and those corresponding to 30°c and 50°c were noted.

After calibration, both vessels were placed on solid carbon dioxide for two days. They were then allowed to reach thermal equilibrium at 0°c; the level of the meniscus in both cases indicated 100% conversion to grey tin.

After one hour in melting ice for the estimation of the degree of transformation, the dilatometers were moved to a bath at 20°c for a further hour as a control measure before being placed in the appropriate thermostat for transformation at 30 or 50°c. Readings of the meniscus level were taken at 30 second intervals (30°c) or 15 second intervals (50°c), until the transformation rates had fallen to a low value. The initial movement of the meniscus is due to: (i) the $\alpha \rightarrow \beta$ transformation, leading to a falling value; (ii) thermal expansion, leading to a rising value. It was hoped that, by conducting control experiments with each specimen in the white modification, thermal expansion curves could be subtracted from the experimental curves to obtain the transformation characteristics. Experience showed that this procedure is satisfactory for transformation at 30°c, since little transformation occurred in the five minutes required for the dilatometers to come to thermal equilibrium with their surroundings; a small anomalous maximum at the beginning of the (transformation, time) curve is, however, observed, owing to the smaller thermal conductivity of grey tin producing a reduced heat exchange between the bath and the dilatometer during transformation tests than during the establishment of the correction curve using white tin. The resultant thermal lag is more important during transformation at 50°c since the reaction usually neared completion after eight minutes; results at 50°c will therefore be described qualitatively only.

(i) *Tests at* 50°c

Specimens were repeatedly transformed between successive tests; 23 hours at the temperature of solid carbon dioxide were allowed for the $\beta \rightarrow \alpha$ change, and 20 to 30 minutes at 50°c was sufficient for forming white tin. Tests were made after 1, 3 and 15 transformations. The anomalous maximum in the curve of the percentage of grey tin against time fell as each specimen was thermally cycled, corresponding to an increase in transformation velocity as found by previous workers (Cohen and van Lieshout 1935 a). For D1, the maxima in the experimental curves appeared after 80 seconds in the thermostat, and the magnitudes showed that, at this point, the amount of transformed grey tin had increased by approximately 10% between the second and third tests. With specimen D2 the maxima occurred 100 seconds after the beginning of the reaction, with an increase of approximately 5% between these two tests. Transformation was complete, judging from observed volume changes, after 30 minutes for specimen D1, but for specimen D2, though completion was approached in a similar period after one and three transformations, the experimental curve in the final test flattened at 81% transformation after 80 minutes, no further change taking place during the next hour.

During the repeated transformations the two specimens appeared identical in the grey state, forming a loose powdery mass with little tendency for individual particles to adhere. On conversion to white tin at 50°c, the specimen in ethyl alcohol gradually changed to an agglomerated mass, while D1 remained powdery.

(ii) *Tests at* 30°c

Similar tests were carried out for each specimen at 30°c, after 5, 17 and 33 transformations. Dilation curves with and without transformation were obtained and subtraction of the thermal expansion correction from the curves involving transformation resulted in the (transformation, time) curves shown in figures 1 and 2. These show clearly that the reaction rate for the specimen under xylene increases with repeated transformation, while that for the specimen under alcohol falls. The maximum transformation rate in each test (the steepest tangent to the curve) is given in table 2.

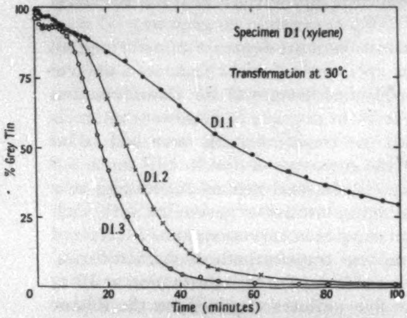


Figure 1. Percentage transformation as a function of time at 30°c, using xylene as dilatometric liquid ($\alpha \rightarrow \beta$).

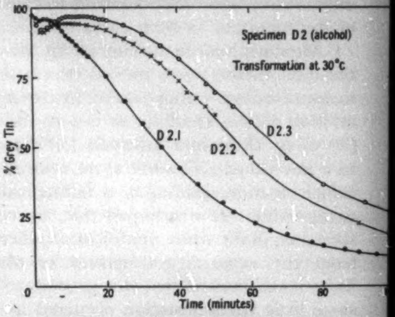


Figure 2. Percentage transformation as a function of time at 30°c, using alcohol as dilatometric liquid ($\alpha \rightarrow \beta$).

Table 2

(1)	(2)	(3)	(4)	(5)	(6)	(7)
D1·1	D1	5	0·58	32	74	13
2	D1	17	1·95	20	68	1
3	D1	33	2·84	15	65	1
D2·1	D2	5	0·94	25	67	4
2	D2	17	0·80	54	55	2
3	D2	33	0·68	63	55	2

(1) Test ; (2) specimen ; (3) number of previous changes ; (4) maximum rate $\alpha \rightarrow \beta$ (g min^{-1}) ; (5) time for maximum to occur (min) ; (6) % α present at maximum rate ; (7) % α present after 180 minutes.

When tests D1.1 and D2.1 were stopped after 1000 minutes, the amount of retained grey tin was 5% for specimen D1, but transformation was complete in specimen D2. All other tests were discontinued after 180 minutes. The quantity of white tin present after 24 hours on solid carbon dioxide following transformation was in all cases less than 1% in D1 and undetectable in D2.

§ 3. DISCUSSION

(*a*) Table 1 shows that the liquid in which dilatometric experiments on the tin transformation are carried out has a significant effect on transformation rates for grey tin not previously subjected to thermal cycling. The velocity of transformation increases with temperature for all specimens, a much higher value being obtained at 29·86°C than at 25·60°C. Other authors have suggested that a maximum rate may be observed. Thus Cohen and van Lieshout (1935 a), using a dilatometric liquid described as xylol, obtained a maximum transformation rate at 25·60°C, while Dunkerley and Munge (1950) found this maximum to occur at 23°C, using iso-amyl alcohol.

In tests 1 and 2 (table 1) specimen LS2 showed the highest activity. At 29·86°C specimen LS4 appeared more active than LS2 after six minutes. When, however, the experimental (dilation, time) curves were plotted, the gradient decreased more markedly for LS2 than for LS4. This may be interpreted as due to a greater activity in the case of LS2, resulting in a much higher degree of transformation in the six minutes before the initial observation. Thus the experimental curve for LS2 was showing the decay of the reaction, while that for LS4 represented an earlier stage in the process.

(*b*) The specimen transformed in pink salt solution (LS3) possessed the lowest $\alpha \rightarrow \beta$ rates, but showed a relatively high $\beta \rightarrow \alpha$ rate at 1·48°C. Groen (1954) reports that almost any type of tin ion will accelerate the $\beta \rightarrow \alpha$ change in pure tin. In the present work, however, transformation did not take place when Vulcan white tin, covered with alcoholic pink salt solution, was maintained at sub-zero temperatures in the absence of grey tin nuclei and electro-replaceable impurities. Cohen and van Eijk (1899, 1900), Tamman and Dreyer (1931) and Kamar and Lasarew (1935) have suggested that the action of pink salt is due to electrolytic cells of the type white tin – tin ions – grey tin, resulting in the local deposition of grey tin at temperatures below the transition point. Difficulties are raised however by the semiconducting properties of grey tin, and by Groen's observation that the rate of transformation at a white tin – grey tin boundary, where the concentration of such local cells would be high, is similar to that elsewhere on the sample. A more satisfactory interpretation may be based on the description of a white tin particle, formed from grey tin, put forward by Dunkerley and Munge (1950). These authors suggest that a coherent rigid shell of white tin surrounds a small nucleus of grey tin, complete transformation of which at temperatures above the critical temperature is prevented by the strain energy accompanying further volume decrease within the rigid shell. Below the critical temperature, transformation will first take place in the interior of the particle and the growing volume of grey tin will ultimately† reach the surface of a given particle. Provided that no barriers exist, such as oxide films, the grey tin may nucleate the transformation on the surface of a contiguous particle, of different size, thus giving rise to an enhanced transformation rate. The role of the pink salt solution is then, by its reducing character, to prevent the formation of oxide films on the white tin surfaces and thus allow nucleation from particle to particle to occur. At the same time, the growth of grey tin within particles of white tin will cause surface stresses of varying magnitude, and local cells may be set up between neighbouring white tin surface areas of different strain energy. At the cathodic areas, in the alcoholic pink salt medium, ions will be deposited either as highly unstable white tin which immediately transforms, or directly as grey tin. In either event centres of inoculation are formed, tending to accelerate the transformation rate.

(*c*) The change of $\alpha \rightarrow \beta$ transformation rate on repeated transformation is of interest in view of the results in table 1. In this table, the specimen under alcohol has a much higher transformation rate at 29·86°c than that under xylene. Similar behaviour is shown (see table 2) in specimens D1 and D2 after five successive transformations. On further repeated transformation, however, the transformation rate at 30°c increases forD 1 (xylene) but decreases for D2 (alcohol). At 50°c an increasing transformation rate was observed on repeated transformation of both specimens.

In seeking to interpret these results, significance is attached to the coherence of white tin particles (but not grey tin) in the alcohol medium, while no such coalescence was observed in xylene. Alcohol thus appears to exert an appreciable 'bonding' action on white tin, whereas xylene, a hydrocarbon, does not. If this is accepted, interpretation of the transformation rates may be attempted as follows.

When a singly transformed grey tin specimen is raised to a temperature above the transformation point, white tin forms at isolated and randomly distributed points on the surfaces of particles. The cohesion of white tin particles in alcohol suggests that, in alcohol, mutually attractive forces may occur between neighbouring white tin areas, tending to compress the surface layers between them; this corresponds to compression of the particle, and an enhanced transformation rate since the change to white tin causes a reduction in volume. In xylene, the cohesive effect is absent and the lower transformation rate may be understood.

On repeated transformation, the activity (the potential ability to transform) increases, owing to a reduction in particle size and an increase in the strain energy content. The initial rate of white tin formation thus increases at 30°c, and large white tin areas are rapidly formed; these, because of the volume reduction, are likely to be severely fissured. If the cohesive effect of alcohol is to prevent such fissures from opening, and thus form a relatively coherent and rigid shell of white tin, the initial compression in the interior of the particle changes, as transformation proceeds in the interior, to a tensional stress which is unfavourable to further transformation, the rate of which may be expected to decrease. Since, however, the white tin shell is not completely rigid, gradual collapse of the white tin upon the residue of grey tin allows the reaction to approach completion in time. With xylene, no interference with the tendency for increased transformation rates on repeated transformation would be expected. The continued decrease of rate in alcohol with the number of thermal cycles may be attributed to the progressive diminution of particle size and the occurrence of finer, more abundant, fissures in the white tin initially formed.

---

† The growth of grey tin within a given particle will cause compressional constraints on the grey phase, the magnitudes of which will depend on particle size. Thus grey tin will not reach the surface of all particles simultaneously, assuming a range of particle sizes.

Similar considerations may also explain the high $\alpha \rightarrow \beta$ velocity at 29·86°c in water (table 1), since an –OH group is present in this case also. To account for the increase in transformation rate at 50°c under both xylene and alcohol, it must be assumed that the higher transformation potential at the higher temperature overcomes the restraining influences of the cohesive forces resulting from the alcoholic medium. Some restraint remains, as shown by the smaller increase for D2 than for D1 in the amount of material transformed after a given time at 50°c between the second and third tests.

The similar linear rates of growth of grey tin in wires, in spite of the variation of liquid medium, may be understood, since the liquid did not make direct contact with the white tin, and growth was across a single interface devoid of oxide films.

In view of the significant role of the liquid medium in dilatometric experiments on the tin transformation, results obtained by this method and reported as characteristic of pure tin must be accepted with caution. Thus the maximum $\alpha \rightarrow \beta$ rate at 23°c reported by Dunkerley and Munge may be due to the cohesive effects of iso-amyl alcohol becoming more appreciable above this temperature, since in this temperature range a relatively large number of centres of white tin would be expected to form on any given particle. Also the material had undergone at least three transformations before the dilatometric experiment. It is probable that their results are true only for the particular medium used, and do not represent an intrinsic property of tin. The results obtained by Cohen and van Lieshout (1935 b) may be similarly qualified.

(*d*) It is of interest to discuss tentatively the origin of the cohesive effect exerted on white tin by alcohol. If a film of alcohol is placed on the oxide-free surface of a metal, coordination of the oxygen of the hydroxyl group to unsaturated surface atoms may be possible, giving rise to extensive wetting. In this case the walls of fine fissures would tend to be drawn together, whereas with a hydrocarbon like xylene, no such action would be expected. According to this viewpoint, the presence of a hydroxyl group in the liquid medium is expected to cause similar effects to those described for alcohol, but to extents depending on the precise molecular structure of the liquid.

§ 4. CONCLUSIONS

The experiments described demonstrate that the liquid medium employed affects the rate of transformation of tin in dilatometric experiments. The actual transition temperature is, however, independent of the liquid used. In view of the effect of liquids, the results of previous workers, in the absence of very detailed experimental data, cannot necessarily be generalized beyond the conditions of their experiments. Further work is necessary to investigate the physical nature of the effect of the medium, and experiments using highly pure tin and a wider range of liquid media are envisaged. It would also be of interest to investigate the effects of liquids on the tensile properties of white tin.

ACKNOWLEDGMENTS

The work described forms part of a wider programme on the transformation of tin, which has been in progress in the University of Birmingham, England, and is being extended in the Metallurgy Department of Toronto University, Canada. Grateful acknowledgment is made to the Tin Research Institute for sponsoring one of us (R. W. S.) and for financial assistance, which, together with that from the Imperial Chemical Industries Ltd., has greatly assisted the work.

REFERENCES

COHEN, E., and VAN LIESHOUT, A. K. W. A., 1935 a, *Z. Phys. Chem.*, **173**, 1 ; 1935 b, *Ibid.*, **173**, 32.

COHEN, E., and VAN EIJK, C., 1899, *Z. Phys. Chem.*, **30**, 601 ; 1900, *Ibid.*, **33**, 59.

DUNKERLEY, F. J., and MUNGE, W. L., 1950, *University of Pennsylvania, Technical Report*, January.

GROEN, L. J., 1954, *Proc. Acad. Sci. Amst.* B, **57**, 122.

KOMAR, H., and LASAREV, B., 1935, *Phys. Z. Sowjet.*, **7**, 468.

TAMMAN, G., and DREYER, K. L., 1931, *Z. anorg. Chem.*, **199**, 97.

Lattice Vibrations and Harmonic Forces in Solids

By A. J. E. FOREMAN and W. M. LOMER
Atomic Energy Research Establishment, Harwell, Berks.

MS. received 2nd July 1957

Abstract. A discussion is given of the interpretation in terms of harmonic forces between atoms of the phonon frequency against wave vector data which is provided by the results of both x-ray and neutron scattering experiments. A knowledge of the phonon polarization vectors is required before a calculation of the components of the interatomic harmonic force tensors between all neighbour atom pairs can be made, but for lattice waves propagating along the symmetry directions of a crystal the Fourier series expansions of the dispersion curves give directly the harmonic force constants between entire planes of atoms. The interpretation of these interplanar force constants in terms of a model with interatomic tensor harmonic forces between a limited number of atom neighbours may require that the interplanar force constants satisfy an internal consistency relation ; the failure of the x-ray data for aluminium to satisfy this relation exactly suggests either that atoms beyond the fourth nearest neighbour are interacting appreciably or that the concept of static harmonic forces between atoms is breaking down.

§ 1. INTRODUCTION

EXPERIMENTS on both the diffuse scattering of x-rays and the inelastic scattering of neutrons by the lattice vibrations (phonons) in a metallic single crystal can provide information on the frequency of a lattice vibrational wave as a function of the magnitude and direction of its wave propagation vector. This information may in turn be used to determine the details of the interatomic forces acting on the metal atoms when they suffer small displacements from their equilibrium positions, i.e. the harmonic forces between atoms, provided that we everywhere assume that the vibrational motion of the atoms in a metallic crystal may be adequately described on the basis of the adiabatic approximation.

Investigations of this type have been made using x-rays for aluminium (Walker 1956), copper (Jacobsen 1955), zinc (Joynson 1954) and α-iron (Curien 1952 a, b), and some preliminary neutron diffraction studies have also been made for aluminium by Brockhouse and Stewart (1955) and Carter, Palevsky and Hughes (1957). The results of such experimental work can usually be shown to be compatible with a model of a crystal in which each atom interacts with only a limited number of its neighbouring atoms, but there is not in general a very precise quantitative agreement with highly simplified models, e.g. those having only central forces between nearest neighbour atoms, even when some allowance has been made for the additional volume dependent terms that arise from the contribution of the conduction electrons to the crystal elasticity (Horton and Schiff 1956). The force constants of aluminium have also been estimated by Salter (1955) from the measured values for the elastic constants and high temperature behaviour of the specific heat and calorimetric entropy, assuming interaction forces between first and second neighbour atoms.

A detailed calculation of the interatomic harmonic force constants, or even of the elastic constants, of a metal crystal from electron theory presents great difficulties since our knowledge of the electron wave functions and cohesive forces in a metal is as yet incomplete (Mott 1956, Löwdin 1956, Jaswon 1954). Some approximate calculations of this type have, however, been attempted recently by Brenig (1955).

In the present paper we shall discuss the interpretation of the results of the x-ray and neutron scattering experiments in terms of the interatomic and interplanar harmonic force constants. A knowledge of these harmonic force constants is of value in providing a starting point for a calculation of the lattice specific heat of a solid (Blackman 1955). In addition, they may be of use for investigating the structure and properties of point and line defects (e.g. vacancies, interstitial atoms, dislocations) in a crystal lattice, since the harmonic force model can provide a rather more detailed description of a lattice strain field than does the elastic continuum approximation (Kanzaki 1957, Maradudin 1956).

§ 2. LATTICE VIBRATION THEORY

The general theory of lattice vibrations in crystals has been described in detail by several authors, e.g. Born and Huang (1954), Peierls (1956), Seitz (1940), and will be only briefly outlined here†. These theories are based on two approximations. First, the *adiabatic approximation* is used to postulate that the total energy of the solid is the sum of the kinetic energy of the atoms and a potential energy depending only on the instantaneous nuclear coordinates. This is tantamount to the assumption that the electrons are always able to adapt themselves completely to the instantaneous nuclear positions (Seitz 1940). It is not, however, immediately obvious from the electron theory of metals (Peierls 1956) that such an approximation is in general valid in metals, and we shall therefore bear in mind in the present paper the possibility of using the experimental data to test directly the validity of the adiabatic approximation. Then, if the adiabatic approximation is made the potential energy of the lattice may be written in a general Taylor series in terms of the displacements $\mathbf{u}_n$ of the nth atom from its equilibrium position $\mathbf{a}_n$, so that

$$U-U_0=\tfrac{1}{2}\sum_{n,n'}\mathbf{A}_{n,n'}\mathbf{u}_n\mathbf{u}_{n'}+\tfrac{1}{6}\sum_{n,n'n''}\mathbf{B}_{n,n',n''}\mathbf{u}_n\mathbf{u}_{n'}\mathbf{u}_{n''}+\dots \qquad \dots\dots(1)$$

where for simplicity we have restricted the discussion to the case of crystals with one atom per unit cell.

For small displacements only the leading quadratic term is considered, i.e. the *harmonic approximation*. Though the other terms are small they are responsible for thermal resistance and thermal expansion. Thus we write

$$U-U_0=\tfrac{1}{2}\sum_{n,n'}\mathbf{A}(\mathbf{a}_n-\mathbf{a}_{n'})\mathbf{u}_n\mathbf{u}_{n'}, \qquad \dots\dots(2)$$

---

† In the present text we shall use the notation adopted by Peierls (1956).

so that the force on the nth atom is

$$\mathbf{F}_n = -\sum_{n'} \mathbf{A}(\mathbf{a}_n - \mathbf{a}_{n'})\mathbf{u}_{n'}, \qquad \ldots\ldots(3)$$

where $\mathbf{A}(\mathbf{a}_n)$ is the harmonic force tensor between atoms connected by a lattice vector $\mathbf{a}_n$ (see also Laval 1956). It is a symmetric tensor if each atom lies at a centre of symmetry and therefore contains six independent components, which number will be further reduced if $\mathbf{a}_n$ is a symmetry direction of the crystal.

The equations of motion yield plane wave solutions of the form

$$\mathbf{u}_n = \mathbf{v}(\mathbf{f}) \exp\,[i\mathbf{f}\,.\,\mathbf{a}_n - i\omega(\mathbf{f})t],$$

where the vector $\mathbf{v}(\mathbf{f})$ and frequency $\omega(\mathbf{f})$ satisfy

$$m\omega^2\mathbf{v} = \mathbf{G}(\mathbf{f})\mathbf{v} \qquad \ldots\ldots(4)$$

in which $\mathbf{G}(\mathbf{f})$ is a tensor given by

$$\mathbf{G}(\mathbf{f}) = \sum_n \mathbf{A}(\mathbf{a}_n) \exp\,(i\mathbf{f}\,.\,\mathbf{a}_n). \qquad \ldots\ldots(5)$$

The vibrational frequencies are the three roots of the secular equation

$$|\mathbf{G}(\mathbf{f}) - m\omega^2\mathbf{I}| = 0 \qquad \ldots\ldots(6)$$

and are denoted by $\omega(\mathbf{f}, s)$ $(s = 1, 2, 3)$, and the corresponding eigenvector by $\mathbf{v}(\mathbf{f}, s)$.

On transforming equation (5) we have that

$$\mathbf{A}(\mathbf{a}_n) = \frac{1}{N} \sum_{\mathbf{f}} \mathbf{G}(\mathbf{f}) \exp\,(-i\mathbf{f}\,.\,\mathbf{a}_n), \qquad \ldots\ldots(7)$$

where the summation is over vectors $\mathbf{f}$ in the first Brillouin zone of the reciprocal lattice and may be replaced by an integration in the limit of large N. Equation (4) may be rewritten in the form

$$\mathbf{V}(\mathbf{f})\boldsymbol{\Omega}(\mathbf{f}) = \mathbf{G}(\mathbf{f})\mathbf{V}(\mathbf{f}), \qquad \ldots\ldots(8)$$

where $\mathbf{V}(\mathbf{f})$ is the matrix formed by compounding the three column vectors $\mathbf{v}(\mathbf{f}, 1), \mathbf{v}(\mathbf{f}, 2), \mathbf{v}(\mathbf{f}, 3)$ into a 3×3 matrix, and $\boldsymbol{\Omega}(\mathbf{f})$ is a 3×3 diagonal matrix with elements $m\omega^2(\mathbf{f}, 1), m\omega^2(\mathbf{f}, 2), m\omega^2(\mathbf{f}, 3)$. Thus from equations (7) and (8) the harmonic force tensor is given by

$$\mathbf{A}(\mathbf{a}_n) = \frac{1}{N} \sum_{\mathbf{f}} \mathbf{V}(\mathbf{f})\boldsymbol{\Omega}(\mathbf{f})\mathbf{V}^{-1}(\mathbf{f}) \exp\,(-i\mathbf{f}\,.\,\mathbf{a}_n). \qquad \ldots\ldots(9)$$

It will be seen from equation (9), that a knowledge of both the phonon frequency $\omega(\mathbf{f}, s)$ and the polarization vector $\mathbf{v}(\mathbf{f}, s)$ (for the three branches $s = 1, 2, 3$) is required before a calculation of $\mathbf{A}(\mathbf{a}_n)$ can be carried out.

It is not in general possible to calculate the harmonic force tensor $\mathbf{A}(\mathbf{a}_n)$ if only the phonon frequency $\omega(\mathbf{f}, s)$ is known, but from the properties of matrices

$$\text{Trace } \mathbf{A}(\mathbf{a}_n) = \frac{1}{N} \sum_{\mathbf{f}} \left[\sum_{s=1}^{3} m\omega^2(\mathbf{f}, s) \right] \exp\,(-i\mathbf{f}\,.\,\mathbf{a}_n), \qquad \ldots\ldots(10)$$

and therefore the *trace* of the harmonic force tensor may be found from a knowledge of $\omega(\mathbf{f}, s)$ without requiring the phonon polarization vector†. It should be noticed, however, that whilst the trace of the harmonic force tensor will contain useful information regarding the nature of the interatomic forces, it may not reflect all of their characteristic properties. If we suppose, for example, that the potential energy is a sum of two-body terms

---

† The following relations exist between the elements $G_{\alpha\beta}$ of a 3×3 matrix and its eigenvalues λ_s $(s = 1, 2, 3)$:

(i) $\text{Trace } G_{\alpha\beta} = \lambda_1 + \lambda_2 + \lambda_3,$

(ii) $\begin{vmatrix} G_{11} G_{12} \\ G_{21} G_{22} \end{vmatrix} + \begin{vmatrix} G_{22} G_{23} \\ G_{32} G_{33} \end{vmatrix} + \begin{vmatrix} G_{33} G_{31} \\ G_{13} G_{11} \end{vmatrix} = \lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_2\lambda_1,$

(iii) $\text{Det. } G_{\alpha\beta} = \lambda_1\lambda_2\lambda_3,$

where λ_s denotes $m\omega^2(\mathbf{f}, s)$ in the present context. Only the linear trace relation may be simply related to the harmonic force tensor, however, although the remaining two relations contain additional information on the form of $\mathbf{G}(\mathbf{f})$.

$$U=\tfrac{1}{2}\sum_{n,\,n'}\phi(\mathbf{r}_n-\mathbf{r}_{n'}), \qquad \ldots\ldots(11)$$

where $\mathbf{r}_n=\mathbf{a}_n+\mathbf{u}_n$ is the position of the nth atom after displacement, then it follows that

$$\text{Trace } \mathbf{A}(\mathbf{a}_n)=[\nabla^2\phi(\mathbf{r}_n)]_{\mathbf{r}_n=\mathbf{a}_n}, \qquad \ldots\ldots(12)$$

and therefore the trace of the harmonic force tensor vanishes dentically for all two-body interatomic potentials which satisfy Laplace's equation. Thus, although it is tempting to assume that the decrease of the trace of $\mathbf{A}(\mathbf{a}_n)$ with increasing $|\mathbf{a}_n|$ correctly reflects the decrease with distance of all the elements of $\mathbf{A}$, i.e. that we can determine the range of the interatomic forces from the trace relation, this assumption could be seriously in error. In particular ϕ might be an unscreened Coulomb potential, for which the trace of $\mathbf{A}(\mathbf{a}_n)$ is zero, but for such a potential the individual elements of the force tensor do not vanish. If, however, ϕ has some other form, such as a screened Coulomb potential or a Lennard–Jones potential, for example, then the decrease of the trace with increasing distance will correctly represent the general behaviour of all the elements of the force tensor.

§ 3. Phonons Propagating Along Crystal Symmetry Directions

Experiments on the diffuse scattering of x-rays by metal single crystals (see references given in § 1) give the details of $\omega(\mathbf{f},s)$ for the longitudinal and transverse lattice waves propagating along the symmetry directions of a crystal, e.g. the ⟨100⟩, ⟨110⟩, and ⟨111⟩ directions in a face-centred cubic crystal, and it is therefore pertinent to discuss here the interpretation of these dispersion curves in terms of the interatomic harmonic force constants.

When the wave propagation vector $\mathbf{f}$ lies along one of the symmetry directions of a cubic crystal it may be shown that the secular equation (6) for $\omega(\mathbf{f})$ factorizes and may therefore be solved directly. The physical interpretation of this is not difficult to see since lattice waves propagating along a symmetry direction correspond to the vibrations of complete *planes* of atoms, so that the mathematical problem is now one of a linear chain of particles, each connected to its neighbour by harmonic forces, and where each particle represents a plane of atoms. Furthermore, it is not difficult to show that as a result of the symmetry of the atoms in the {100}, {110} and {111} planes of a face-centred cubic crystal the longitudinal and two transverse modes of relative displacement are independent of one another for each type of plane, so that each longitudinal and transverse mode of vibration may be treated separately as a linear chain problem.

The vibrational frequency of a linear chain of particles, each having only one degree of freedom, is given by (cf. equation (4))

$$m\omega^2(f)=2\sum_{n=0}^{\infty}P(n)\cos nfa, \qquad \ldots\ldots(13)$$

where a is the particle spacing and $P(n)$ is the harmonic force constant between

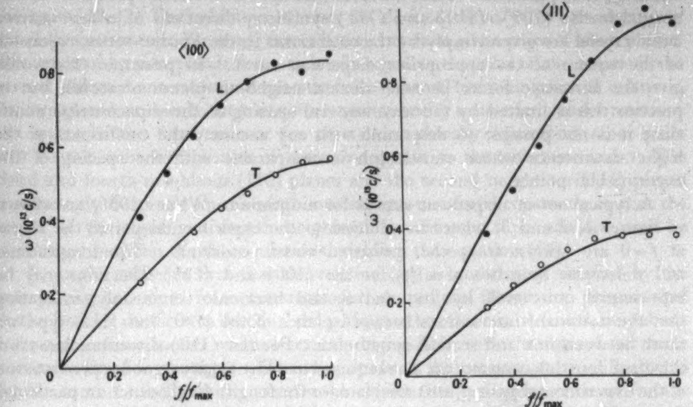


Figure 1. Figure 2.

Figures 1 and 2. X-ray data for elastic waves propagating along the ⟨100⟩ and ⟨111⟩ directions in aluminium (Walker 1956). The smooth curves correspond to the first two harmonic terms in the Fourier series expansion of $\omega^2(f)$, with coefficients chosen to give the best fit to the elastic constants and x-ray data.

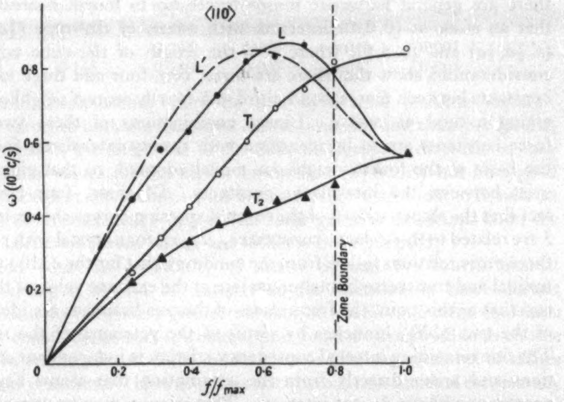


Figure 3. X-ray data for elastic waves propagating along the ⟨110⟩ direction in aluminium (Walker 1956). The smooth curves correspond to the first four harmonic terms in the Fourier series expansion of $\omega^2(f)$, with coefficients chosen to give the best fit. The broken curve L′ denotes the closest longitudinal dispersion curve which is consistent with the other six dispersion curves on the basis of a general fourth neighbour model.

*n*th neighbour particles. Thus the force constants between the planes of atoms† normal to the ⟨100⟩, ⟨110⟩ and ⟨111⟩ symmetry directions of a face-centred cubic crystal are given simply by the coefficients in the Fourier series expansion of the square of the appropriate dispersion curve. In principle this would give the harmonic forces between distant neighbour planes of atoms, but in practice this is limited by the accuracy and spacing of the experimental points since it is not possible to determine with any accuracy the coefficients of the higher harmonics whose wavelength is comparable with the spacing of the experimental points.

A typical set of dispersion curves for aluminium (Walker 1956) are shown in figures 1, 2 and 3, where in addition to the experimental points the slopes at $f=0$ are known from the measured elastic constants. The longitudinal and transverse branches of $\omega^2(f)$ for the ⟨100⟩ and ⟨111⟩ directions may be represented quite well by first and second harmonic terms only, indicating that the dominant interactions between planes of the {100} and {111} type are those between first and second neighbours. For the ⟨110⟩ direction, however, third and fourth harmonics are also required in order to give a good representation of the experimental points (see table) and for the longitudinal branch in particular the fourth harmonic term is quite appreciable ; this suggests that the fourth neighbour harmonic forces are not negligible, which is not in agreement with the conclusion arrived at by Walker that only first and second neighbour interatomic forces are important in aluminium.

In discussing the harmonic forces between planes of atoms it is important to consider in detail their relation to the interatomic harmonic forces. We shall consider in particular a model of the face-centred cubic lattice in which there are general harmonic tensor forces up to fourth nearest neighbours, so that an atom at $(0,0,0)$ interacts with atoms of the type $(\frac{1}{2}a, \frac{1}{2}a, 0)$, $(a, 0, 0)$, $(a, \frac{1}{2}a, \frac{1}{2}a)$ and $(a, a, 0)$, where a is the length of the cube edge. Symmetry considerations show that there are three, two, four and three independent force constants between first, second, third and fourth nearest neighbours respectively, giving a total of twelve. Linear combinations of these twelve interatomic force constants are to be identified with twenty interplanar force constants on the basis of the fourth neighbour model adopted, so that eight relations must exist between the interplanar constants. Of these, four correspond to the fact that the slopes at $\mathbf{f}=0$ of the seven dispersion curves shown in figures 1, 2 and 3 are related to three elastic constants c_{11}, c_{12}, c_{44} for a crystal with cubic symmetry ; three more relations follow from the condition that for the ⟨110⟩ direction a longitudinal and transverse branch must join at the extreme value of the wave vector $\mathbf{f}$ and that at this point the frequencies of the two branches are identical with those of the two ⟨100⟩ branches by virtue of the geometry of the reciprocal lattice. The one remaining internal consistency relation is independent of such considerations and arises directly from the assumption that atoms beyond the fourth nearest neighbour do not interact. This relation may be expressed in the form

$$3\,\mathrm{trace}\,\mathbf{P}^{(3)}(110) + 12\,\mathrm{Trace}\,\mathbf{P}^{(4)}(110) = 4\,\mathrm{Trace}\,\mathbf{P}^{(2)}(111), \qquad \ldots\ldots(14)$$

where $\mathbf{P}^{(n)}(\alpha\beta\gamma)$ denotes the harmonic force tensor between *n*th neighbour $\{\alpha\beta\gamma\}$ planes and the trace of this tensor is simply the sum of the longitudinal and two transverse harmonic force constants between *n*th neighbour planes of the type $\{\alpha\beta\gamma\}$. The relation (14) may be derived directly by considering in detail the interatomic harmonic forces acting between given pairs of atom planes. The trace of an interplanar force tensor is equal to the sum of the traces of the several different interatomic harmonic force tensors between pairs of atoms in the two planes, and since the trace of a tensor quantity is invariant under rotations the traces of the force tensors for crystallographically equivalent atom pairs will be the same irrespective of their orientation to the reference axes. The third and fourth neighbour (110) planes and the second neighbour (111) planes involve only third and fourth neighbour interatomic force constants, so that the traces of the third and fourth neighbour interatomic force tensors may be eliminated between the three relations, yielding the equation (14).

---

† It is convenient to define P(n) as the harmonic force per atom acting between *n*th neighbour planes.

From the table we find that the relation (14) is not closely satisfied by the interplanar harmonic force constants for aluminium (these are proportional to the Fourier coefficients listed in the table), so that there is no fourth neighbour interatomic force model which is exactly compatible with the interplanar force constants listed in the table. The magnitude of the inconsistency is illustrated by the curve L′ in figure 3 ; this is the closest fitting dispersion curve for the longitudinal ⟨110⟩ lattice waves which is compatible with the other six curves on the basis of the fourth neighbour model.

	a_0	a_1	a_2	a_3	a_4
L(100)	0·386	−0·352	−0·034	—	—
T(100)	0·151	−0·157	0·006	—	—
L(111)	0·448	−0·470	0·022	—	—
T(111)	0·080	−0·080	0·000	—	—
L′(110)	0·417	−0·190	−0·262	0·033	0·002
L(110)	0·408	−0·254	−0·205	0·097	−0·046
T_1(110)	0·330	−0·365	0·029	0·013	−0·007
T_2(110)	0·145	−0·140	0·007	−0·017	0·005

This apparent inconsistency may be due to several possible factors : (i) experimental errors in both the measured elastic constants and the x-ray results, and including errors in the latter due to the difficult Compton corrections, may be giving anomalous values for the higher harmonics, (ii) it may not be a good approximation to neglect interactions with atoms further than the fourth nearest neighbours, (iii) the inconsistency may reflect some genuine deviations from the model of static harmonic forces which has hitherto been adopted.

In connection with the possibility of atoms beyond the fourth nearest neighbour interacting it should be noted that the inclusion of more distant neighbour harmonic forces gives a greater number of interatomic force constants than interplanar constants, so that it would appear to be always possible to choose the parameters in an interatomic force model so as to give any given set of values for the interplanar constants. However, it is not obvious that internal consistency relations do not still exist for a model having interactions with more distant neighbours, and we find in fact that a relation of the same type as that described above still exists for a model with interactions up to the eighth nearest neighbour atoms.

§ 4. SUMMARY

It is shown that a knowledge of the branches of the phonon frequency $\omega(\mathbf{f}, s)$ for all values of $\mathbf{f}$ in the first Brillouin zone of the reciprocal lattice is only sufficient to determine exactly the *traces* of the harmonic force tensors between all neighbour atom pairs. For phonons propagating along certain symmetry directions of a crystal, where the polarization vectors are known, the $\omega(\mathbf{f}, s)$ data yield directly the harmonic force constants between entire *planes* of atoms, but there is some difficulty in choosing a set of interatomic harmonic force constants which correspond precisely to those interplanar constants.

The symmetry of the reciprocal lattice imposes certain relations between the interplanar force constants. If we attempt to relate the interplanar constants to interatomic forces between atoms out to fourth neighbours, we find an additional restriction on the interplanar constants, which is not well satisfied by the x-ray data for aluminium. The discrepancy may be due to experimental inaccuracies, to anharmonic terms in the potential and breakdown of the adiabatic approximation, or to the presence of longer range interatomic forces than those assumed.

Polarization vector determination by neutron diffraction is difficult and inaccurate. Detailed information on the phonon frequencies for symmetry directions will be more useful for interpretation in terms of interatomic forces than results for general points in reciprocal space. Frequency determinations at such points can be used only to derive some idea of the range of forces through the trace relation. The interplanar forces, which are derived directly from the symmetry direction results, give just as clear an indication of this range.

REFERENCES

BLACKMAN, M., 1955, *Handbuch der Physik*, **7**, pt. 1, 325 (Berlin : Springer).

BORN, M., and HUANG, K., 1954, *Dynamical Theory of Crystal Lattices* (Oxford : University Press).

BRENIG, W., 1955, *Z. Phys.*, **142**, 163.

BROCKHOUSE, B. N., and STEWART, A. T., 1955, *Phys. Rev.*, **100**, 756.

CARTER, R. S., PALEVSKY, H., and HUGHES, D. J., 1957, *Phys. Rev.*, **106**, 1168.

CURIEN, H., 1952 a, *Bull. Soc. Franç. Minér.*, **75**, 197 ; 1952 b, *Acta Cryst., Camb.*, **5**, 392.

HORTON, G. K., and SCHIFF, H., 1956, *Phys. Rev.*, **104**, 32.

JACOBSEN, E. H., 1955, *Phys. Rev.*, **97**, 654.

JASWON, M. A., 1954, *The Theory of Cohesion* (London : Pergamon Press).

JOYNSON, R. E., 1954, *Phys. Rev.*, **94**, 851.

KANZAKI, H., 1957, *J. Phys. Chem. Solids*, **2**, 24.

LAVAL, J., 1956, *C.R. Acad. Sci., Paris*, **242**, 2502.

LÖWDIN, P., 1956, *Advanc. Phys.* (*Phil. Mag.* Suppl.), **5**, 1.

MARADUDIN, A. A., 1956, *Ph.D. Thesis*, University of Bristol.

MOTT, N. F., 1956, *Nature, Lond.*, **178**, 1205.

PEIERLS, R. E., 1956, *Quantum Theory of Solids* (Oxford : University Press).

SALTER, L., 1955, *Proc. Roy. Soc.* A, **233**, 418.

SEITZ, F., 1940, *The Modern Theory of Solids* (New York : McGraw-Hill).

WALKER, C. B., 1956, *Phys. Rev.*, **103**, 547.

Stacking Faults and Dislocations in Copper–Aluminium and Cobalt–Nickel Alloys

By J. W. CHRISTIAN and J. SPREADBOROUGH
Department of Metallurgy, The University Museum, Oxford

MS. received 22nd July 1957

Abstract. Stacking fault densities α' in heavily cold-worked powders of copper–aluminium and cobalt–nickel alloys have been measured from the peak shifts, using Paterson's theory. The Fourier coefficients of the broadening function due to faults are calculated for the {111} and {200} powder lines, assuming equal densities of faults on all {111} planes, and these enable the strain broadening function to be evaluated from the measured coefficients. This function has approximately a Cauchy shape. Dislocation densities ρ estimated from the strain breadth ξ and mean square strain $\langle S^2\rangle$, vary much less with composition than does α'. Faults and dislocations disappear together on annealing, so that there is a linear relation between ρ and α'. It is nevertheless suggested that the stacking fault diffraction effects cannot be attributed to the fault ribbons in extended dislocations, as previously stated.

---

§ 1. Introduction and Experimental Work

Deformation stacking faults in measurable amounts have recently been found in heavily worked powders of certain cubic and hexagonal alloys (Warren and Warekois 1955, Anantharaman and Christian 1956, Christian and Spreadborough 1956, Smallman and Westmacott 1957). For the cubic materials, the shifts in the positions of the diffraction peaks are in accordance with the theory due to Paterson (1952), but the physical interpretation of the results is uncertain. In particular, it is not known whether the observed effects may be ascribed to the fault ribbons present as extended dislocations, since a diffraction theory for this situation is not available. The experiments described below were designed to measure the stacking fault parameter α and the Fourier coefficients of the strain broadening function. The results indicate that for a given alloy, α is proportional to the dislocation density.

Copper–aluminium alloys were used for the main investigation, since resistivity measurements suggest that this system is very suitable for stacking fault work. Similar measurements on these alloys have recently been reported by Smallman and Westmacott (1957), but their method of analysis differs from that given here. The analysis was also extended to some cobalt–nickel alloys, for which values of α were already known.

The alloys were made homogeneous under identical conditions at 950°c. Filings from the ingots were used to prepare diffractometer specimens, and profiles of the {111}, {200}, {220} and {311} powder lines were obtained by hand scanning, using copper Kα radiation and a Norelco diffractometer. The low angle 'tail' of the {111} line was measured with especial care, at least 6400 counts being taken at each angular interval to ensure adequate statistical accuracy. The specimens were subsequently annealed, and the profiles redetermined.

The Kα_1 profiles were separated from the composite lines graphically (Rachinger 1948) and by an analytical procedure (Anantharaman and Christian 1953) ; the results of the two methods were in good agreement. The Kα_1 lines were found to be symmetrical about their peak positions, to within the limits of the experimental accuracy. The intensities of the {111} and {200} lines were tabulated at angular intervals corresponding to two-hundredths of the complete period (defined below). The Fourier coefficients of the diffraction broadening were evaluated by Stokes method (Stokes 1948), using the profiles of the completely annealed powders as standards.

§ 2. LINE SHAPE ANALYSIS

The use of Stokes' method when the diffraction broadening produces a shift in the peak position was considered by Garrod and Auld (1955) for strain broadening, but has otherwise been rather neglected. Suppose two functions separately transform a unit of intensity at $x=0$ into distributions $f(x)$, $g(x)$. When both effects are present, the resultant distribution is given by the convolution transform

$$h(x)=\int g(u)f(x-u)\,du.$$

Now take new origins at $x=d_1$, d_2, so that $f(x)=a(x-d_1)=a(x_1)$ and $g(x)=b(x-d_2)=b(x_2)$, and write $h(x)=c(x-d_1-d_2)=c(x_3)$. The resultant is then

$$c(x_3)=\int b(u)a(x_3-u)\,du.$$

If the functions $a(x_1)$, $b(x_2)$ and $c(x_3)$ are written as Fourier series over the ranges in which x_1, x_2 and x_3 respectively vary from $-\frac{1}{2}p$ to $+\frac{1}{2}p$, the coefficients of these series are related by $A(n)B(n)=C(n)/p$.

In applying Stokes' method, the usual tacit assumption is made that if $f(x)$ and $g(x)$ have maxima at $x=0$, $h(x)$ also has a maximum there. The corresponding more general result would be that if the maxima are at d_1 and d_2, the maximum of $h(x)$ is at d_1+d_2. In fact, these statements are valid if origins are taken at the mean values of the distributions, but not for maxima, unless the line shapes are symmetrical, when the mean values and maxima coincide. Fortunately, this is the case for most types of broadening encountered in practice, so that Fourier coefficients may be calculated with respect to origins at the peak positions.

The diffraction broadening from a deformed face-centred cubic structure may be due to internal strains, particle size effects and stacking faults. These are all manifestations of the number and distribution of dislocations and other defects, and are separated in this way mainly for convenience of analysis. In the treatment by Warren and Averbach (1950, 1952 a, b), it is assumed that particle size and strain broadening produce a diffraction profile which is symmetrical about the Bragg angle for the undeformed structure, and may be expressed as

$$a(x)=\sum_{-\infty}^{\infty}A(n)\cos 2\pi nx \qquad \ldots\ldots(1)$$

where x is a linear displacement from the powder line peak. The function $a(x)$ is thus defined between limits $x=\pm\frac{1}{2}$, which must be chosen so that the intensity is effectively zero outside them.

Measurement of deformation faults in face-centred cubic materials has previously been based on Paterson's theory, assuming that the peak shifts for the faults on different {111} planes are linearly additive, but more detailed analysis is required if line profiles are to be considered. The resultant powder profile may be predicted if all four sets of {111} planes contain an equal density of faults, a not unreasonable assumption after very heavy deformation. The distribution is obtained by calculating separately the variation of intensity with Bragg angle for each set of faults, and using convolution transforms to find the resultant, or by calculating the three dimensional intensity distribution round a reciprocal lattice point, and transforming this into the Bragg angle variation. The first method is rather easier to apply, providing the Bragg angle variation has the same period for all the sets of faults, as is the case for the {111} and {200} lines. For the {111} line, the two methods give the same answer when the usual approximation is used for transferring the reciprocal lattice distribution into the powder distribution. In calculating the powder profiles, it is only necessary to consider one component of each line, since the peak shifts and profiles of all (*hkl*) components of a given {*hkl*} powder line are identical when there are faults on all the {111} planes.

Let the faults on *each* set of {111} planes be specified by Paterson's parameter α, and let $\sigma=(1-3\alpha+3\alpha^2)^{1/2}$. (The symbol σ is used instead of ρ to avoid confusion with the dislocation density.) The distribution in each line may then be written

$$b(x_r)=\sum_{-\infty}^{\infty} B_r(n)\cos 2\pi n x_r \qquad \ldots\ldots(2)$$

where the subscript r identifies the line ($r=111$, 200 etc.), and x_r is a measure of the angular distance from the (displaced) peak. The line peak is at a Bragg angle θ_r, its displacement from the peak (θ_0) of the undistorted structure is $\Delta\theta_r$, and θ is the Bragg angle at any point in the profile. Introducing the quantity $y=[1/6-(1/2\pi)\tan^{-1}\sqrt{3}(1-2\alpha)]$, the explicit results required for the {111} and (200) lines in the present work may be written

$$\left.\begin{array}{ll}\Delta\theta_{111}=3\lambda^2 y/2a^2\sin 2\theta & B_{111}(n)=\sigma^{3n}\\ \quad x_{111}=(2a^2\sin 2\theta/\lambda^2)(\theta-\theta_{111}) & \\ \Delta\theta_{200}=-4\lambda^2 y/a^2\sin 2\theta & B_{200}(n)=\sigma^{4n}\\ \quad x_{200}=(a^2\sin 2\theta/\lambda^2)(\theta-\theta_{200}) & \end{array}\right\} \qquad \ldots\ldots(3)$$

where a is the side of the cubic unit cell, and λ the x-ray wavelength.

The measured diffraction profile, after correction for instrumental broadening, is the convolute of (1) and (2) and can be put in the form

$$c(x_r)=\sum_{-\infty}^{\infty} C_r(n)\cos 2\pi n x_r \qquad \ldots\ldots(4)$$

where $C_r(n)=A(n)B_r(n)$, and x in (1) has been made equal to x_r. The derivation assumes that Paterson's theory adequately describes the broadening, as well as the peak shifts. From the peak shifts, $B_r(n)$ may be calculated and used to obtain $A(n)$ from the measured $C(n)$. Since x_r is fixed, the period used cannot be arbitrary, and Bevers–Lipson strips are not suitable for the summations.

Paterson suggested that the stacking fault parameter should be determined from a relation between the real and imaginary parts of each complex Fourier coefficient, measured with respect to an origin at the Bragg angle of the peak position in the undistorted structure. This, however, does not provide an independent measure of α, since it is equivalent to measuring the peak displacement. Apart from the consistency of α values from different lines, the only experimental test of the validity of the theory is to examine the symmetry of the line about its peak. If no asymmetry can be detected, the equations for $\Delta\theta_r$ and Paterson's equation (10) *must* give the same result for α.

§ 3. RESULTS

Figure 1 shows the variation of stacking fault parameter with composition for the copper–aluminium alloys ; in order to facilitate comparison with previous results, the quantity plotted is the total stacking fault density $\alpha' = 4\alpha$. The curve has the same shape as those previously found for copper–zinc and cobalt–nickel alloys (Warren and Warekois 1955, Christian and Spreadborough 1956), and the previously noted correlation with the resistivity change is also valid for this system. Some results obtained by Smallman and Westmacott (1957) are also included in figure 1. The value $\alpha' = 0{\cdot}051$ obtained for the 8% alloy is the highest yet recorded for face-centred cubic deformation faulting.

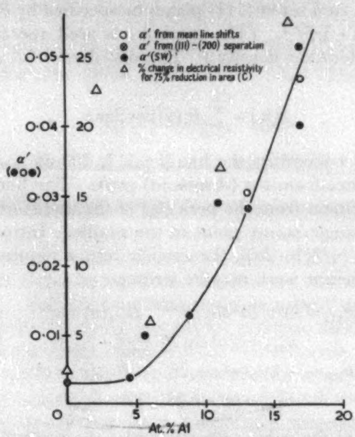


Figure 1. Stacking faults in copper–aluminium alloys. SW, Smallman and Westmacott, C, Crampton *et al.*

The significant contribution which stacking faults make to the observed diffraction broadening is illustrated in figure 2, which shows the measured Fourier coefficients $C(n)$ and the coefficients $A(n)$ obtained from them. In Warren and Averbach's analysis of strain broadening, it is shown that if the variable x of equation (1) is expressed as $x = 2a_3(\sin\theta - \sin\theta_0)/\lambda$, the Fourier coefficients are

$$A(n) = A_L = \langle \cos 2\pi(h^2+k^2+l^2)^{1/2}\Delta L/a \rangle$$

where the brackets represent a mean value. The quantity ΔL is the change in length of a column of crystal of original length $L = na_3$ normal to the (hkl) plane.

Since $x = x_r$, in terms of which the measured coefficients $C(n)$ are defined, the lengths L are given by

$$L = na(h^2+k^2+l^2)^{1/2}/(h+k+l).$$

For a Gaussian profile, or if ΔL is very small, the average of the cosine term gives the equation usually used in interpreting the results

$$A_L = \exp\{-2\pi^2(h^2+k^2+l^2)\langle \Delta L^2\rangle/a^2\}. \qquad \ldots\ldots(5)$$

The $(A(n), n)$ relations were converted to $(\langle \Delta L^2\rangle^{1/2}, L)$ plots in this way; an example is shown in figure 3. The use of a period large compared with the measurable part of a line gives points on this curve for smaller values of L than are usually obtained, but these points are inaccurate because of the tail errors. The curve approximates to a parabola, although there is a finite slope at the origin, whilst the corresponding $(\langle \Delta L^2\rangle, L)$ relation is a straight line. Williamson and Smallman (1954) have shown that this result is expected if equation (5) is used to analyse a line profile with a Cauchy rather than a Gaussian

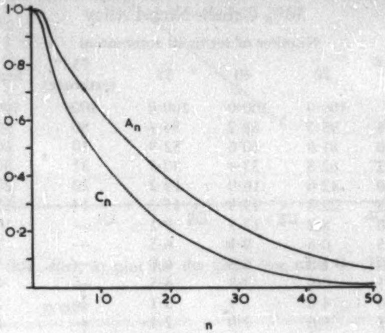


Figure 2. Fourier coefficients $C(n)$ and $A(n) = C(n)/\sigma^{3n}$ for {111} line from Cu–6% Al alloy.

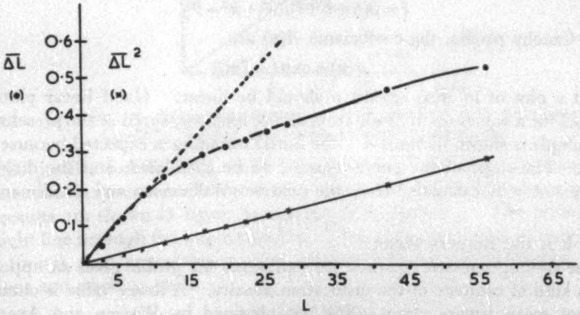


Figure 3. Variation of $\langle \Delta L^2\rangle^{1/2}$ and of $\langle \Delta L^2\rangle$ with L from {111} line in Cu–4% Al alloy.

shape. Bowman, Krumhansl and Stock (1955) have also found the $(\langle \Delta L^2 \rangle, L)$ relations are linear when the results are analysed in this way.

The implication that the strain coefficients $A(n)$ represent a Cauchy profile was tested for the 50% nickel alloy and the 6% aluminium alloy by direct calculation. Table 1 gives comparative values of the intensity obtained by Fourier summation, and from the theoretical expression

$$a(x_r) = (1 + \pi^2 x_r^2/\beta^2)^{-1}$$

where β is the integral breadth in the units of x_r. The agreement is poor when only 20 terms are included in the summation, but is quite good for 40 terms, and still better for 55. From the equation for x_r, it follows that the integral breadth of the strain distribution ξ is given by

Table 1. Reconstruction of Strain Profile, $a(x_r)$, from Coefficients $A(n)$ for 50% Cobalt–Nickel Alloy

	Number of terms in summation				
x^r	20	40	55	75 (estimate)	Eqn (6)
0	100·0	100·0	100·0	100	100·0
0·005	95·3	88·2	84·4	80	80·0
0·010	81·8	60·6	52·9	50	50·0
0·015	62·8	33·4	30·3	31	30·7
0·020	42·0	16·0	19·2	20	20·0
0·025	22·8	13·9	15·3	14	13·8
0·030	8·4	12·7	9·1	—	10·0
0·035	0·6	9·1	6·5	—	7·5
0·040	−1·8	4·6	6·0	—	5·9
0·045	0·6	2·8	4·5	—	4·7
0·050	4·6	3·9	3·1	—	3·8
0·075	1·6	2·0	2·1	—	1·7
0·100	1·0	0·2	0·0	—	1·0

$$\xi = \beta(h+k+l)/(h^2+k^2+l^2).$$

For a Cauchy profile, the coefficients $A(n)$ are

$$A(n) = \exp(-2n\beta)$$

so that a plot of ln $A(n)$ against n should be linear. Good linear plots were obtained for n values up to 25–30 from all the lines measured in the present work. An example is shown in figure 4 ; the initial rounding is expected because of tail errors. The slope of the curve allows ξ to be calculated, and the dislocation density ρ may be calculated from the relation (Williamson and Smallman 1956)

$$\rho_1 = 16\xi^2/b^2 \qquad \ldots\ldots(7)$$

where **b** is the Burgers vector.

The assumption that equation (6) represents the profile gives an upper limit to this kind of estimate of the dislocation density. A lower value is obtained if the root mean square strain $\langle S^2 \rangle^{1/2}$ is obtained by Warren and Averbach's method of taking the initial slope of a $(\langle \Delta L^2 \rangle^{1/2}, L)$ plot. The dislocation density is then given by

$$\rho_2 = 4\pi \langle S^2 \rangle / b^2. \qquad \ldots\ldots(8)$$

This method was much less satisfactory, since the initial slopes were so ill-defined. Both equations (7) and (8) assume a random distribution.

The results on stacking fault parameter and dislocation density are collected in table 2. The most noteworthy feature is that ρ, unlike α, does not vary appreciably with composition in either alloy system. The densities in the copper–aluminium alloys appear to be substantially larger than in pure copper, but this result must be accepted with reserve, since the result for copper is believed to be much less reliable. For reasons discussed below, the results in the table are believed to give relative densities, but not absolute values of ρ.

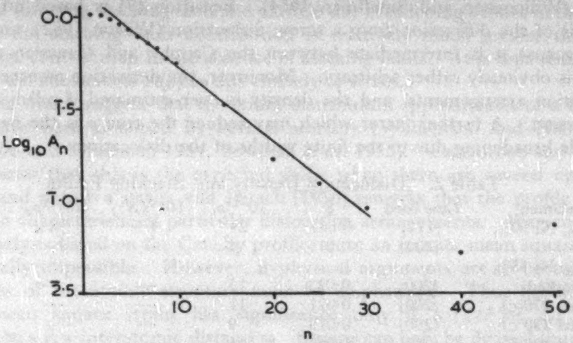


Figure 4. (log $A(n)$, n) plot for the {200} line of a Co–35% Ni alloy.

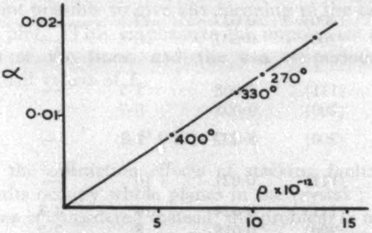


Figure 5. Variation of dislocation density ρ with stacking fault parameter α' for Co–35% Ni alloy after various annealing treatments.

Line profiles for the 35% nickel alloy were measured after short annealing treatments at temperatures below that at which full line sharpness was attained. The results are shown in figure 5 as a plot of α' against ρ. The points lie on a straight line through the origin (the ($\log \alpha'$, $\log \rho$) line has a slope of unity within small limits), thus suggesting a close relation between the formation of dislocations and stacking faults.

§ 4. DISCUSSION

Figure 1 shows some divergence between the present results and those of Smallman and Westmacott. The differences are not much larger than the expected experimental error, some estimate of which may be obtained from the separate plotting of mean α' values determined from the shifts of all four measured lines, and values obtained from the change in the {111} – {200} separation only. The change in separation is believed to give the more accurate result.

Since the shapes are nearly Cauchy, estimates of dislocation density from equation (7) should be more accurate relative to each other than those from equation (8). The absolute values of ρ are, however, very uncertain. For a Cauchy distribution, the stored energy required for the dislocation density estimate can only be obtained by using a cut-off strain. The result ρ of the order of $25\xi^2/b^2$ used by Williamson and Smallman (1956) appears to have been obtained by choosing the strain 10ξ, rather than a fixed value as in their earlier paper (Williamson and Smallman 1954). Equation (7) is based on Wilson's analysis of the diffraction from a screw dislocation (Wilson 1952), and is used also because it is intermediate between the Cauchy and Gaussian estimates, but it is obviously rather arbitrary. Moreover, the derivation assumes random dislocation arrangements, and the density is over-estimated if piled-up groups are present. A further factor which may reduce the true ρ is the neglect of a possible broadening due to the finite widths of the dislocations.

Table 2. Dislocation Density and Stacking Faults

Specimen	Line used	ξ	$10^{-12}\rho_1$	$10^{-4}\langle S^2\rangle$	$10^{-12}\rho_2$	α'
Co 50%, Ni 50%	{111}	0·031	24	4·5	8	0·006
	{200}	0·016	6	—	—	0·006
Co 65%, Ni 35%						
(cold-worked)	{200}	0·023	13	—	—	0·019
(annealed 270°c)	{200}	0·021	11	—	—	0·015
(annealed 330°c)	{200}	0·019	9	—	—	0·013
(annealed 400°c)	{200}	0·016	6	—	—	0·008
Co 69%, Ni 31%						
(cold-worked)	{200}	0·027	17	—	—	0·022
(annealed 500°c)	{200}	0·002	0·1	—	—	0·003
Pure Cu						
	{111}	0·008	1·5	—	—	0·004
	{200}	0·005	0·7	—	—	0·004
(assuming no faults)	{200}	0·007	1·2	—	—	—
Cu 98%, Al 2%	{111}	0·021	10	—	—	0·004
Cu 96%, Al 4%	{111}	0·024	14	4·8	9	0·013
	{200}	0·018	8	2·2	4	0·013
Cu 94%, Al 6%	{111}	0·026	16	—	—	0·031
Cu 92%, Al 8%	{111}	0·029	20	—	—	0·051

The relation $\alpha' = K\rho$, shown in figure 5, lends some support to the hypothesis that measured stacking faults are really extended dislocations. However, a physical argument is given in the Appendix to suggest that Paterson's theory would not apply even approximately to narrow fault ribbons. If this is correct the measured α' values show that appreciable numbers of wide faults are produced during deformation. Some support for this comes from recent electron microscope studies of dislocations in thin foils of austenitic stainless steel (Hirsch 1957, private communication), in which separation of partials to form stacking faults was frequently observed. The theory of diffraction by extended dislocations appears to be very difficult, but it is to be expected that there will be some extra broadening due to the fault ribbon. When this is considered, ρ in table 2 must be reduced. Stored energy measurements confirm, however, that the densities are very high in alloy powders of this type (Hargreaves 1957, private communication).

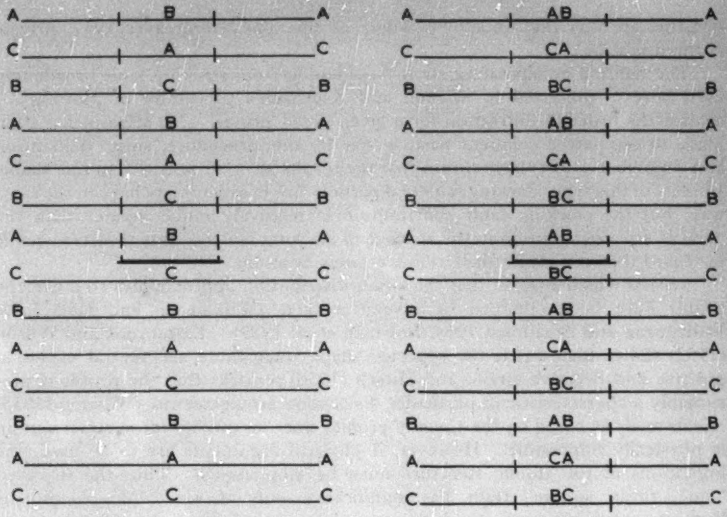
The method of separating strain broadening from stacking fault broadening used here is preferred to Warren and Averbach's procedure of plotting C_L against the order of diffraction for a given set of planes. No attempt has been made to estimate a coherent particle size by this procedure, since Williamson and Smallman (1954) have shown that the results are very sensitive to line shape. In plots of this type, stacking fault and particle size broadening behave in the same way, but the stacking fault contribution is relatively much greater than the particle size contribution in the absence of stacking faults. It is thus reasonable to regard the coefficients $A(n)$ as due entirely to strain.

Results which suggest that the strain distribution approximates to a Cauchy profile have been obtained by several workers (Williamson and Hall 1953, Williamson and Smallman 1954, Bowman *et al.* 1955). Eastabrook and Wilson (1952) shows that this is the expected shape when there are several cycles of positive and negative strain, and Hirsch (1956) remarks that the profile is presumably a characteristic of particular dislocation arrangements. Warren (1955) rejects analysis based on the Cauchy profile, since an infinite mean square strain is physically impossible. However, if physical arguments are to be used, the limitations of the atomic structure must be emphasized. Thus the so-called 'true' mean square strain has significance only if a $(\langle L^2\rangle^{1/2}, L)$ plot is linear up to a few interatomic distances. Strains can only be defined in terms of distances greater than interatomic separations, and unless the strain is constant in this range, it is not possible to give any meaning to the tangent at the origin of the $(\langle L^2\rangle^{1/2}, L)$ plot. This emphasizes the importance of careful measurements of the tails of the lines, and the use of periods sufficiently large to give points for small values of L.

APPENDIX

Calculations of the diffraction effects of stacking faults are made on the assumption that faults occupy whole planes in the crystal ; if a distribution of extended dislocations is considered instead, the problem is much more difficult. At first, it seems attractive to suppose that one can approximately separate two factors, the strain broadening produced by the elastic fields of the half dislocations, and the fault broadening produced by the differences in stacking. The intensity distribution round a reciprocal lattice point due to the second factor might then be given by the function representing large faults, multiplied by the shape transform of the appropriate regions effectively covered by small faults. The following argument seems to show that this is incorrect, and that it is impossible to consider the effect of the fault without including the strain field of the half dislocations.

Figure 6 (*a*) shows a conventional view of the stacking introduced by a pair of Shockley half dislocations, the A, B, C notation signifying the three translational positions of the {111} layers. No elastic relaxation has been allowed, so that the change in stacking at the half dislocations is assumed to lead to a discontinuity on each successive layer in the upper half of the crystal. In addition, the change in stacking is asymmetrical ; this is remedied in figure 6 (*b*) where the notation (AB) etc. now means a layer which is half way between the A and B positions. Note that the dislocations are still half dislocations, and the


A B A
C A C
B C B
A B A
C A C
B C B
A B A
C C C
B B B
A A A
C C C
B B B
A A A
C C C
A AB A
C CA C
B BC B
A AB A
C CA C
B BC B
A AB A
C BC C
B AB B
A CA A
C BC C
B AB B
A CA A
C BC C

(*a*) (*b*)

Figure 6. (*a*) Conventional picture of extended dislocation, with partial slip of upper half of crystal. (*b*) Extended dislocation with stacking changes symmetrical about slip plane.

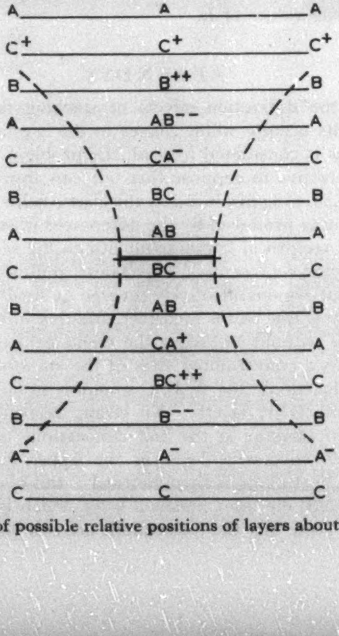

A A A
C^+ C^+ C^+
B B^{++} B
A AB^{--} A
C CA^- C
B BC B
A AB A
C BC C
B AB B
A CA^+ A
C BC^{++} C
B B^{--} B
A^- A^- A^-
C C C

Figure 7. Illustration of possible relative positions of layers about an extended dislocation.

stacking sequence is everywhere . . . . . $\Delta\Delta\Delta\Delta$ . . . . . except at the fault where it is . . . . . . $\Delta\Delta\nabla\Delta\Delta$ . . . . . .

It is clear that neither of these figures can represent the true stacking. A more probable model is shown in figure 7, where the discontinuities disappear in regions remote from the dislocation. The important difference with this model is that for any vector $\mathbf{r}_m$ traversing the stacking fault, the probabilities that two planes m apart are in 1, Δ or ∇ relationship to each other are identical with the same probabilities in the perfect crystal, so long as m is reasonably large. This is quite contrary to the situation assumed in standard stacking fault theory, where all these probabilities must tend to the value 1/3 (independent of m) as $\mathbf{r}_m$ becomes large. The change of stacking introduced by the fault in figure 7 is thus purely localized, whereas with a large fault the stacking of the whole crystal is affected. This is rendered physically plausible by the argument that for large distances, the extended dislocations appear to be normal lattice dislocations, and the effects of the half dislocations merge.

Figure 7 suggests that stacking fault ribbons will not produce equivalent diffraction effects to those calculated for large faults, and it is rather doubtful whether there will be a shift in peak position. A detailed treatment of the problem appears to be extremely difficult.

ACKNOWLEDGMENT

We should like to thank Dr. W. Hume-Rothery, for his interest and encouragement.

REFERENCES

ANANTHARAMAN, T. R., and CHRISTIAN, J. W., 1953, *Brit. J. Appl. Phys.*, **4**, 155 ; 1956, *Acta Cryst., Camb.*, **9**, 479.

BOWMAN, J. C., KRUMHANSL, J. A., and STOCK, J. R., 1955, *J. Appl. Phys.*, **26**, 1057.

CHRISTIAN, J. W., and SPREADBOROUGH, J., 1956, *Phil. Mag.*, **1**, 1069.

CRAMPTON, D. K., BURGHOFF, H. L., and STACEY, J. T., 1941, *Trans. A.I.M.E.*, **143**, 228.

EASTABROOK, J. N., and WILSON, A. J. C., 1952, *Proc. Phys. Soc.* B, **65**, 67.

GARROD, R. I., and AULD, J. H., 1955, *Acta Met.*, **3**, 190.

HIRSCH, P. B., 1956, *Progress in Metal Physics* **6** (London : Pergamon Press) p. 236.

PATERSON, M. S., 1952, *J. Appl. Phys.*, **23**, 805.

RACHINGER, W. A., 1948, *J. Sci. Instrum.*, **25**, 254.

SMALLMAN, R. E., and WESTMACOTT, K. H., 1957, *Phil. Mag.*, **2**, 669.

STOKES, A. R., 1948, *Proc. Phys. Soc.*, **61**, 382.

WARREN, B. E., 1955, *Acta Cryst., Camb.*, **8**, 483.

WARREN, B. E., and AVERBACH, B. L., 1950, *J. Appl. Phys.*, **21**, 595 ; 1952 a, *Ibid.*, **23**, 497 ; 1952 b, *Imperfections in Nearly Perfect Crystals* (New York : Wiley), p. 152.

WARREN, B. E., and WAREKOIS, E. P., 1955, *Acta Met.*, **3**, 473.

WILSON, A. J. C., 1952, *Acta Cryst., Camb.*, **5**, 318.

WILLIAMSON, G. K., and HALL, W. H., 1953, *Acta Met.*, **1**, 22.

WILLIAMSON, G. K., and SMALLMAN, R. E., 1954, *Acta Cryst., Camb.*, **7**, 574 ; 1956, *Phil. Mag.*, **1**, 34.

Geometrical-optical Treatment of Frequency Response

By H. H. HOPKINS
Department of Physics, Imperial College, London

MS. received 19*th June* 1957

Abstract. An expression is obtained for the frequency response of an optical system according to the approximation of geometrical optics. The difficulties inherent in the explicit calculation of ray densities in the image are avoided. The method is applied to defect of focus, astigmatism and aberration tolerances, and the results compared with those obtained according to diffraction theory. In the cases studied the ray-optically calculated frequency response suggests a poorer image quality than that given by diffraction theory.

---

§ 1. Introduction

If the wave front aberration of an optical system is denoted by $W(X, Y)$, the ray having pupil coordinates (X, Y) intersects the image plane in a point having coordinates

$$\xi = -\frac{D}{n}\frac{\partial W}{\partial X}, \quad \eta = -\frac{D}{n}\frac{\partial W}{\partial Y} \qquad \ldots\ldots(1)$$

where D is the radius of the reference sphere, whose centre is at the origin in the plane (ξ, η), and n is the refractive index of the image space. In the approximation of geometrical optics the light intensity at any point is taken to be the ray density at that point, and interference effects are ignored. If we consider initially the special case in which rays from the periphery of a square $\delta x \delta y$ in the pupil intersect the image plane on the periphery of a square $\delta\xi\delta\eta$, the ray density over this element of the image plane will be given, after differentiating equations (1), by

$$\{\tau(X, Y)\}^2 \frac{\delta X \delta Y}{\delta\xi\delta\eta} = \{\tau(X, Y)\}^2 \left(\frac{n}{D}\right)^2 \left\{\frac{\partial^2 W}{\partial X^2}\frac{\partial^2 W}{\partial Y^2}\right\}^{-1} \qquad \ldots\ldots(2)$$

where $\tau(X, Y)$ is the amplitude over the wave front at the point (X, Y). $\{\tau(X, Y)\}^2$ is, of course, the intensity at the point (X, Y), but it is preferable to use the amplitude here for convenience when comparing the resulting formulae with those obtained according to physical optics.

The formula (2) discloses one of the difficulties inherent in treating the distribution of light in star-images by ray optics. At points in the image plane for which either W_x'', W_y'', or both, are zero, the ray density is infinite, and δ-functions must be used to represent the light intensities at these singularities. Points for which either $W_x'' = 0$ or $W_y'' = 0$ will occur when an astigmatic line-focus of the wave element at (X, Y) lies in the chosen image plane; similarly a stigmatic point-focus lies in the image plane when $W_x'' = W_y'' = 0$. A second complication arises in all but the simplest cases, namely that a given element of area in the image plane may receive light from two or more parts of the pupil. These difficulties may be avoided to some extent by the use of spot-diagrams (Linfoot 1955) which give a visual display of the light distribution; but these diagrams do not readily lend themselves to analytical considerations needed for the assessment of image quality.

The frequency response of an optical system may be expressed as the (inverse) Fourier transform of the distribution of intensity in the star-image (Duffieux 1946). Since a δ-function transforms into a constant, we should expect the singularities in the ray-optical light distribution not to appear in the corresponding expression for the frequency response. Hopkins (1955) and De (1955) have given such formulae for the cases of defect of focus and astigmatism respectively. These results were obtained from the calculated ray densities, which take simple forms in these two cases. In what follows we shall show that an explicit calculation of the ray optical light distribution may be avoided; and there results a simple formula for the frequency response according to geometrical optics.

§ 2. RAY-OPTICAL FORMULA FOR FREQUENCY RESPONSE

Let the rays from the periphery of an element of area δA, situated at the point (X, Y) of the pupil, intersect the image plane on the periphery of an element of area $\delta\chi$, situated at (ξ, η). We now place no restriction on the shapes of δA, $\delta\chi$. If $\tau(X, Y)$ is the amplitude at (X, Y), the element of area δA will give rise to a distribution of light δG given by

$$\begin{aligned} \delta G &= \{\tau(X, Y)\}^2 \delta A/\delta\chi && \text{within } \delta\chi \\ &= 0 && \text{outside } \delta\chi \qquad \ldots\ldots(3) \end{aligned}$$

and the total distribution of light $G(\xi, \eta)$ in the image plane will be given by the sum of all such distributions, arising from the different elements δA of the pupil. Since the inverse transform of the sum of a number of terms is equal to the sum of the inverse transforms of the separate terms, we may find the frequency response by adding the inverse transforms of the partial light distributions δG arising from each separate element δA of the pupil. The inverse transform of (3), is

$$\delta g(R_x, R_y) = \{\tau(X, Y)\}^2 \frac{\delta A}{\delta\chi} \iint_{\delta\chi} \exp\{-2\pi i(R_x\xi' + R_y\eta')\}\, d\xi'\, d\eta'$$

where (ξ', η') are coordinates in the image plane, and (R_x, R_y) are spatial frequencies. We may write the above expression as

$$\begin{aligned} \delta g(R_x, R_y) = \{\tau(X, Y)\}^2 &\frac{\delta A}{\delta\chi} \exp\{-2\pi i(R_x\xi + R_y\eta)\} \\ &\times \iint_{\delta\chi} \{1 - 2\pi i[R_x(\xi' - \xi) + R_y(\eta' - \eta)] + \mathrm{O}(\delta\chi^2)\}\, d\xi'\, d\eta' \end{aligned}$$

the variation of ξ' and η' being restricted to the interior of $\delta\chi$. After integration this expression will have the form

$$\delta g(R_x, R_y) = \{\tau(X, Y)\}^2 \delta A \exp\{-2\pi i(R_x\xi + R_y\eta)\} + \mathrm{O}(\delta\chi).$$

In the limit $\delta A \to 0$, the element $\delta\chi \to 0$, and integration over the whole pupil A then leads to the expression

$$g(R_x, R_y) = \iint_A \{\tau(X, Y)\}^2 \exp\{-2\pi i(R_x\xi + R_y\eta)\}\, dA \quad \ldots\ldots(4)$$

for the inverse Fourier transform of the ray-optical light distribution in the image.† (ξ, η) are the coordinates of the point of intersection with the image plane of the ray which passes through the point (X, Y) of the pupil. They are thus the components of the transverse aberration of that ray, and are given as functions of (X, Y) by the relations (1). The normalized frequency response is then found from the formula

---

† *Note added in proof.* Essentially this formula has since appeared in a paper in Japanese (Miyamoto 1957).

$$D(R_x, R_y) = g(R_x, R_y)/g(0, 0) \qquad \ldots\ldots(5)$$

which ensures that $D(0, 0) = 1$.

It is usually more convenient to express the two-dimensional frequency response in terms of single spatial frequencies associated with line structures having different orientations (Hopkins 1955, 1957). For example, the Fourier component (R_x, R_y) has the same frequency response as a component of spatial frequency $R = (R_x^2 + R_y^2)^{1/2}$ of a line structure inclined at an angle $\psi = \tan^{-1}(R_y/R_x)$ to the axis of y. In place of $D(R_x, R_y)$ we shall then write $D(R, \psi)$. If the axes of Y and η are chosen to be parallel with the direction of the line structure, in place of (5) we shall employ the notation

$$D(R, \psi) = g(R, 0)/g(0, 0), \qquad \ldots\ldots(6)$$

R_y being put equal to zero in (4), the value of ξ being that in the new coordinate system. This gives, in place of (4)

$$g(R, 0) = \iint_A \{\tau(X, Y)\}^2 \exp\{-2\pi i R\xi\}\, dA. \qquad \ldots\ldots(7)$$

If (ξ_0, η_0) are the components of the transverse aberration of a ray measured with respect to the original axes, the ξ-component with respect to the new axes (ξ, η) will be given by

$$\xi = \xi_0 \cos\psi - \eta_0 \sin\psi,$$

ψ being the angle between the ξ_0-axis and the direction of the structure for which the frequency response is to be calculated. In terms of the wave front aberration, the components of the transverse ray aberration (ξ_0, η_0) will be determined by the relations (1), but using the wave front aberration $W_0(x_0, y_0)$ specified with respect to pupil axes (x_0, y_0) parallel with (ξ_0, η_0). These axes will usually lie in the sagitta and meridian planes of the optical system.

To facilitate comparison with the frequency response calculated according to wave-optics, we shall introduce the reduced coordinates $x = X/h$, $y = Y/h$ for the pupil, and the reduced variable $s = R\lambda/n \sin\alpha$ for spatial frequencies. The incidence height h is that at which the ray of angular aperture α intersects the reference sphere. If the pupil is circular, and of angular semi-aperture α, the periphery of the pupil is then the circle $x^2 + y^2 = 1$. In place of (1) we now write $(n \sin\alpha)\xi = -W_x'(x, y)$, so that $(-R\xi) = ksW_x'(x, y)$, and the expression (7) is replaced by

$$g(s, 0) = \iint_A \{\tau(x, y)\}^2 \exp\{iksW_x'(x, y)\}\, dx\, dy \qquad \ldots\ldots(8)$$

where $k = 2\pi/\lambda$, the product ks being then independent of the wavelength λ. The normalized frequency response (6) is written

$$D(s, \psi) = g(s, 0)/g(0, 0). \qquad \ldots\ldots(9)$$

In the above formulae the same functional notation is conveniently used despite the changed variables.

§ 3. COMPARISON WITH DIFFRACTION THEORY

To find the frequency response according to diffraction theory, (8) has to be replaced by the formula (Hopkins 1956)

$$g(s, 0) = \iint_S \tau(x + \tfrac{1}{2}s, y)\, \tau(x - \tfrac{1}{2}s, y) \exp\{ikW(x, y; s)\}\, dx\, dy \quad \ldots\ldots(10)$$

where, in place of the area of the whole pupil A, the integration is restricted to the region S, which is that common to two pupils centred on the points $(+\frac{1}{2}s, 0)$ and $(-\frac{1}{2}s, 0)$ respectively. The normalizing factor, $g(0, 0)$, in (9) is unchanged, but the argument of the exponential function in (8) is replaced by

$$W(x,y;s) = W_x'(x,y) + \frac{1}{3!}\left(\frac{s}{2}\right)^2 W_x'''(x,y) + \frac{1}{5!}\left(\frac{s}{2}\right)^4 W_x^{\mathrm{v}}(x,y) + \dots \qquad \dots\dots(11)$$

the various terms containing as factors even powers of s and the successive odd derivatives of the wave front aberration.

It is often suggested that, in problems of image formation, geometrical optics gives a better approximation when the aberrations are large. Comparison of (8) and (10) shows that this is not strictly so in the present case. For aberration affects only the arguments of the exponential functions in (8) and (10), and in these the geometrical optical approximation neglects all terms in (11) other than the first. But the ratios of the moduli of successive terms in (11), for a given s, are unaltered if a given aberration $W(x, y)$ is changed to $pW(x, y)$, p being a numerical factor. What is more, the larger the numerical value of the aberration, the greater will be the absolute values of those terms in (11) which are neglected in geometrical optics. In view of this, one would expect geometrical optics to give not a better, but an inferior, approximation in cases where the aberration is larger.

For a given s the ratios of successive terms in (11) are greater the higher the power of the pupil variables which occur in $W(x, y)$. We should, therefore, expect geometrical optics to give a poorer approximation in the presence of higher order aberrations. The lowest orders of aberration are defects of focus and combined field curvature and astigmatism. For these two cases the wave front aberration has the forms

$$W(x,y) = w_{20}(x^2+y^2)$$
$$W(x,y) = w_{20}(x^2+y^2) + w_{22}(x\sin\psi + y\cos\psi)^2$$

respectively, the axes of x and y making an angle ψ with the principal sections. In both cases the third and higher derivatives of $W(x, y)$ are identically zero, so that no error is committed in making the approximation $W(x, y; s) = W_x'(x, y)$. For this reason an earlier suggestion, that geometrical optics might be reasonably valid for spatial frequencies for which $s < 0{\cdot}10$ (Hopkins 1955), should be treated with some reserve, because this tentative conclusion was based on a comparison of the numerical results of ray and diffraction treatments of pure defect of focus. Nevertheless, the contention that geometrical optics may be used when only low spatial frequencies are in question, rather than for large aberrations, is supported by the present work, as is easily seen from a comparison of (8) and (10) above. If the pupil is of uniform amplitude, $\tau(x, y) = 1$ and the approximation

$$\tau(x+\tfrac{1}{2}s, y)\tau(x-\tfrac{1}{2}s, y) = \{\tau(x, y)\}^2$$

leads to no error. Geometrical optics would, therefore, be expected to be more nearly valid for pupils having uniform transparency.

In the absence of aberrations, $W_x'(x, y) = 0$ and (8), (9) then show the response according to geometrical optics to be equal to unity for all frequencies. Again, according to geometrical optics, the slope $(\partial/\partial s)\,|D(s)|$ is zero at $s = 0$, no matter what aberration is present. In these respects the geometrical optical response differs from that obtained using diffraction theory. On the other hand, it is easily seen that these two results are valid, even according to diffraction theory, for the relative response, which is defined by the relation

$$M(s,\psi)\exp\{i\theta(s,\psi)\}=D(s,\psi)/D_0(s,\psi)$$

in which $M(s,\psi)$ is a modulus, equal to the ratio of the contrast obtained with and without aberration respectively, and the argument $\theta(s,\psi)$ is a phase shift equal to that associated with $D(s,\psi)$ itself. $D_0(s,\psi)$ is used to denote the response in the absence of aberration.

In the geometrical treatment, since $D_0(s,\psi)=1$, there is no distinction between $D_0(s,\psi)$ and $M(s,\psi)\exp\{i\theta(s,\psi)\}$. The calculation of $D_0(s,\psi)$ using diffraction theory is a simple matter. If this value of $D_0(s,\psi)$ is used, together with the value of the relative response obtained according to geometrical optics, in the relation

$$D(s,\psi)=M(s,\psi)\,D_0(s,\psi)\exp\{i\theta(s,\psi)\} \qquad \ldots\ldots(12)$$

the value of the response obtained might be expected to be more accurate than that given by geometrical optics alone. Results obtained by this mixed treatment are given below.

Lohmann (1957, private communication) has considered the influence of the form of the transparency function of the pupil, $\tau(x,y)$, on the frequency response; the response $D(s,\psi)$ being expanded as a power series in s up to the fourth power. If the geometrical optical formula (8) is considered from this point of view, it is to be observed that the region of integration, the modulus of the exponential function and its argument, which occur in the diffraction result (10), need to be expanded separately and the leading term of each retained. The region of integration and the modulus of the integrand then become independent of s, but the exponential factor retains all powers of s if expressed as a series.

§ 4. APPLICATIONS

The ray-optical frequency response has been treated for the two cases of defect of focus (Hopkins 1955) and astigmatism (De 1955). In each case the response was obtained by Fourier inversion of the calculated distribution of ray density in the images of point sources. The same formulae may be obtained more simply using (8) and (9) above. To evaluate the response in an azimuth ψ in any focal plane in the presence of astigmatism, the wave front aberration is of the form

$$W(x,y)=w_{20}(x^2+y^2)+w_{22}(x\sin\psi+y\cos\psi)^2 \qquad \ldots\ldots(13)$$

in which w_{22} measures the astigmatism and w_{20} the distance of the chosen image plane from the sagittal focus. We shall assume that the wave fronts are of uniform amplitude, so that $\tau(x,y)=1$. From (13)

$$ksW_x'(x,y)=s\{[a+(b-a)\sin^2\psi]x+[(b-a)\sin\psi\cos\psi]y\} \qquad \ldots\ldots(14)$$

where $a=(4\pi/\lambda)w_{20}$ and $b=(4\pi/\lambda)(w_{20}+w_{22})$. Using polar coordinates in place of (x,y), and assuming a circular pupil,

$$D(s,\psi)=\frac{1}{\pi}\int_0^1\int_0^{2\pi}\exp\{irS\cos(\phi-\epsilon)\}r\,dr\,d\phi,$$

ϵ being a constant, and the value of S being given by

$$S=s(a^2\cos^2\psi+b^2\sin^2\psi)^{1/2}. \qquad \ldots\ldots(15)$$

The integration yields immediately

$$D(s,\psi)=\frac{2J_1(S)}{S} \qquad \ldots\ldots(16)$$

which is identical with the result obtained by De (1955). If we put $b=a$, corresponding to there being no astigmatism, $S=as$, and (16) becomes

$$D(s,\psi)=\frac{2J_1(as)}{as} \qquad \ldots\ldots(17)$$

which is the result for pure defect of focus (Hopkins 1955).

The above expressions are to be regarded as giving the values of

$$M(s,\psi)\exp\{i\theta(s,\psi)\}$$

in the 'mixed' treatment. We see that $\theta(s,\psi)=0$ or π in both cases, and the frequency response according to the treatment is given by the formula

$$D(s,\psi)=\frac{2J_1(S)}{S}D_0(s,\psi) \qquad \ldots\ldots(18)$$

in place of (16). For a circular aperture of uniform amplitude $D_0(s,\psi)$ has the well-known form

$$D_0(s,\psi)=\frac{1}{\pi}(2\beta-\sin 2\beta) \quad \beta=\cos^{-1}|\tfrac{1}{2}s| \qquad \ldots\ldots(19)$$

and is independent of azimuth.

A comparison of the results obtained by the geometrical optical treatment and those according to diffraction is shown in figure 1, for the case of defect of focus with a circular pupil. For the two curves $w_{20}=6\lambda/\pi$, $12\lambda/\pi$ respectively,

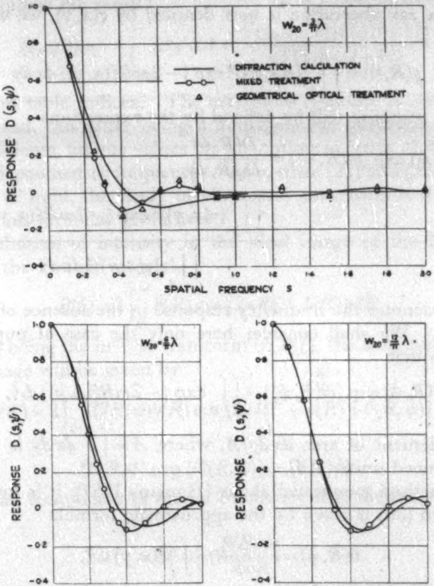


Figure 1. Comparison of frequency response according to diffractions and ray optics.

the difference between the purely geometrical and the mixed treatments is too small to be shown. Any difference between the purely geometrical and the mixed treatments would be expected only for the smaller values of aberration, for only then will $D(s, \psi)$ have values differing significantly from zero when s is large enough for $D_0(s, \psi)$ to be appreciably less than unity. It would appear from these curves that the geometrical-optical treatment could usefully be applied to the present case for the region $|s| < 0{\cdot}10$, with an error of less than 0·08.

The curves of figure 1 show better agreement between geometrical optics and diffraction theory for the larger values of aberration. It should be noted, however, that this arises indirectly, from the fact that it is only for small values of s that the response is of significant magnitude when the aberration is larger, and it is for small values of s that geometrical-optics is more applicable. In this sense, therefore, the use of geometrical optics may be considered as appropriate when the aberrations are large.

One advantge of using frequency-response methods is that it has made possible the formulation of a diffraction treatment of tolerances, valid for both small and large aberrations (Steel 1952, Hopkins 1957). The same method can be applied to the geometrical optical formulae (8) and (9). For this it is found more convenient to use the formulae (6) and (7), rather than (8) and (9), but to retain the fractional pupil variables (x, y). This means the omission of a constant factor from the expression for $g(R, 0)$, which, however, cancels on forming the ratio (6). If the transverse ray aberration is now denoted by $\xi(x, y)$, we write in place of the formula (7)

$$g(R,0) = \iint_A \{\tau(x,y)\}^2 \exp\{-2\pi i R \xi(x,y)\}\, dx\, dy$$

and the relative response will be defined by the expressions

$$M(R,\psi)\exp\{i\theta(R,\psi)\} = \frac{D(R,\psi)}{D_0(R,\psi)} = \frac{\iint_A \{\tau(x,y)\}^2 \exp\{-2\pi i R\xi(x,y)\}\, dx\, dy}{\iint_A \{\tau(x,y)\}^2\, dx\, dy}$$

where $D_0(R, \psi)$ denotes the frequency response in the absence of aberration and defect of focus. We shall consider here only the case of pupils of uniform transparency, so that

$$M(R,\psi)\exp\{i\theta(R,\psi)\} = \iint_A \exp\{-2\pi i R\xi(x,y)\}\, dA, \qquad \ldots\ldots(20)$$

dA being the element of area $dx\,dy/A$, where $A = \iint_A dx\, dy$ is the area of the pupil in the reduced units (x, y).

Using the method mentioned above (Hopkins 1957), it is easily shown that the phase shift in (20) is given by the approximate formula

$$\theta(R,\psi) = (-2\pi R)\iint_A \xi(x,y)\, dA, \qquad \ldots\ldots(21)$$

provided the product $2\pi R\xi(x, y)$ is never large; and also that the modulus of (20) satisfies the inequality

$$M(R,\psi) \geqslant |-2\pi^2 R^2 K_G|, \qquad \ldots\ldots(22)$$

provided the right-hand side is greater than or equal to 0. The factor K_G is given by

$$K_G = \iint_A \{\xi(x,y)\}^2\, dA - \left\{\iint_A \xi(x,y)\, dA\right\}^2 \qquad \ldots\ldots(23)$$

and is the minimum mean square value of the ξ-component of the transverse ray aberration, measured with respect to an η-axis parallel with the direction of the structure considered. Thus, if $\xi_0(x_0, y_0)$, $\eta_0(x_0, y_0)$ are the rectangular components of the transverse aberration, measured with respect to the customary axes, the relation

$$\xi(x,y) = \xi_0(x_0, y_0)\cos\psi - \eta_0(x_0, y_0)\sin\psi$$

will give the aberration component $\xi(x,y)$ to be used for a line structure inclined at an angle ψ to the meridian section y_0.

In the diffraction treatment a less simple expression has to be used in place of $\xi(x,y)$, and, for a circular pupil, the integration extends over a region S which is that common to two circular pupils centred on the points $(\pm\frac{1}{2}s, 0)$. This has necessitated rather extensive tabulation of the integral

$$H_{k,l}(s) = \iint_S x^k y^l\, dS.$$

In the geometrical-optical treatment, on the other hand, the case of a circular pupil merely requires the result

$$H_{k,l}(0) = \iint\limits_{x^2+y^2\leqslant 1} x^k y^l\, dA = \frac{\Gamma(\frac{1}{2}\{k+1\})\Gamma(\frac{1}{2}\{l+1\})}{\pi\Gamma(\frac{1}{2}\{k+l+4\})} \qquad \ldots\ldots(24)$$

and a single short table suffices. The maximum response is obtained when K_G is minimized; and, this latter being a homogeneous quadratic function of the aberration coefficients, unique values for the optimum ratios of these coefficients are obtained independent of frequency. Also, since (21) and (22) do not contain the wavelength of light, the study of chromatic aberration is simplified by the geometrical optical formulae given here.

If the distribution of intensity in the ideal image of the line structure is represented by the Fourier integral,

$$B(\xi) = 2\int_0^\infty b(R)\cos\{2\pi R\xi + \phi(R)\}\, dR,$$

$b(R)\exp\{i\phi(R)\}$ being the inverse transform of $B(\xi)$, the distribution of intensity in the actual image will be given by

$$B'(\xi') = 2\int_0^\infty T(R,\psi)b(R)\cos\{2\pi R\xi' + \phi(R) + \theta(R,\psi)\}\, dR$$

where

$$T(R,\psi)\exp\{i\theta(R,\psi)\} = D(R,\psi).$$

Thus the Fourier component of spatial frequency R is displaced laterally by an amount

$$\delta\xi = -\frac{\theta(R,\psi)}{2\pi R}$$

or, according to the approximate formula (21), by

$$\delta\xi = \iint_A \xi(x,y)\, dA \qquad \ldots\ldots(25)$$

which is independent of R. Now (21) holds with good accuracy provided $M(R, \psi)$ is not less than about 0·8, and the geometrical-optical approximation becomes increasingly more accurate as the frequency tends to zero. Hence the phase shift $\theta(R, \psi)$ may be expected merely to produce a displacement of the image as a whole in a lens which is well corrected for use in conditions in which only the lower frequencies are of importance. The same result has been found to hold in the fuller diffraction treatment of tolerances (Hopkins 1957). The above result (25) shows that, to this accuracy, the image will be centred on the centre of gravity of the ray-intersection pattern. K_G is then seen to represent the square of the radius of gyration of this distribution about an axis passing through the centre of gravity and lying parallel to the direction ψ of the structure to be imaged. A number of authors have considered the radius of gyration of the ray-intersection pattern about an axis perpendicular to the image plane as a measure of image quality based on point source images. If this axis is taken to pass through the centre of gravity of the ray distribution, the resulting radius of gyration will be given by $K^2 = (K_G)_x{}^2 + (K_G)_y{}^2$ and represents a weighted average of K_G with respect to azimuth. It should be noted that K here represents the radius of gyration of the geometrical optical image. Lansraux (1953) has shown that this radius can become infinite for the diffraction image.

The transverse ray aberration in the image plane may be found from the relation

$$\xi(x, y) = -\frac{1}{n' \sin \alpha'}\, W_x{}'(x, y).$$

For primary spherical aberration and defect of focus the wave front aberration has the form

$$W(x, y) = w_{20}(x^2 + y^2) + w_{40}(x^2 + y^2)^2$$

and (23) gives for this case

$$K_G = \left(\frac{1}{n' \sin \alpha'}\right)^2 \iint\limits_{x^2 + y^2 \leqslant 1} \{2w_{20}x + 4w_{40}(x^2 + y^2)x\}^2\, dA$$

or, with the notation of (24),

$$K_G = \left(\frac{1}{n' \sin \alpha'}\right)^2 \{4w_{20}{}^2 H_{2,0}(0) + 16w_{40}{}^2[H_{6,0}(0)$$
$$+ 2H_{4,2}(0) + H_{2,4}(0)] + 16w_{20}w_{40}[H_{4,0}(0) + H_{2,2}(0)]\}.$$

Inserting the numerical values of $H_{k,l}(0)$ gives

$$K_G = \left(\frac{1}{n' \sin \alpha'}\right)^2 \left\{w_{20}{}^2 + 2w_{40}{}^2 + \frac{8}{3}\, w_{20}w_{40}\right\}$$

which has a minimum value

$$K_G = \left(\frac{1}{n' \sin \alpha'}\right)^2 \frac{2w_{40}{}^2}{9} \qquad \ldots\ldots(26)$$

when $w_{20} = -4w_{40}/3$. This condition, which determines the focal plane having the maximum contrast, is identical with that obtained from diffraction theory for the limit $s \rightarrow \psi$. Lohmann (1957, private communication) has determined the optimum focal plane in the presence of spherical aberration for pupils of uniform and non-uniform amplitude, by seeking the plane of focus for which $|g''(0)|$ is minimized, and in this way has obtained the same result for the case of uniform amplitude. This is to be expected, since the approximate formulae (21) and (23) are the initial terms in the power series in R for the two quantities in question. On the other hand the formula (20) contains no approximation other than the assumption that geometrical optics is valid.

If we impose as tolerance criterion the condition $M(s, \psi) \geqslant 0{\cdot}80$, which may be written

$$K_G \leqslant \frac{0{\cdot}10}{\pi^2 R'^2} \qquad \ldots\ldots(27)$$

we find, using (26),

$$|w_{40}| \leqslant \frac{0{\cdot}21(n' \sin \alpha')}{R'} = \frac{0{\cdot}21\lambda}{S} \qquad \ldots\ldots(28)$$

as the tolerance for primary spherical aberration to this order of approximation. For a system of relative aperture F/5, and 40 lines per mm, this tolerance has the value $|w_{40}| = 0{\cdot}5\,\mu$, that is approximately 1λ, which is the value allowed by the Rayleigh tolerance and the Strehl intensity criterion (Definitionshelligkeit). For $R' = 20$ lines per mm, on the other hand, double this tolerance is allowable.

A comparison of the optimum focal plane, and of the tolerance for primary spherical aberration, according to diffraction theory and geometrical optics, is shown in figure 2. The tolerance according to geometrical optics is in all these cases smaller than that given by diffraction theory. Furthermore, at least within the greater part of the central band of frequencies, the curves of figure 1 show that the geometrical-optical approximation leads to a smaller value of response than the diffraction treatment. For systems such as photographic lenses, for which only the smaller spatial frequencies (s less than 0·10–0·20) come in question, the geometrical optical treatment would seem to be useful, both because it is somewhat simpler to apply in practical cases and because the errors appear to be such that an underestimate of the performance is given. This last conclusion is to be expected in one respect, namely because the whole pupil region A is greater than that of the region S, which is that common to two pupils relatively displaced, and it is precisely over the additional outer region that the integrand in (10) will usually be expected to be of greatest variation. For this reason, the mean value of the integrand over the region A will be expected to be less than the

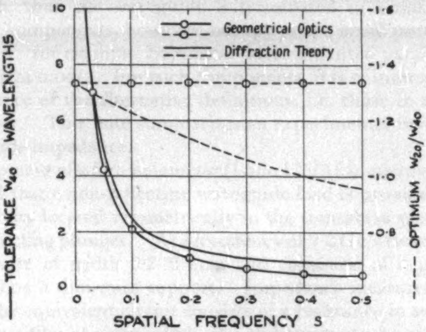


Figure 2. Optimum focus and tolerance for primary spherical aberration.

mean value over the smaller region S. On the other hand, to maximize the response with respect to a variable aberration coefficient implies choosing the ratios of the aberration coefficients so as to maximize the mean value of the integrand in (10) over the region S. For the reason given above, this condition may well differ appreciably from that which maximizes the integrand of (8) over the whole pupil region A. Hence, except for small values of s, the best forms of correction according to geometrical optics may be significantly different from those determined according to diffraction theory.

REFERENCES

De, M., 1955, *Proc. Roy. Soc.* A, **233,** 91.

Duffieux, P. M., *L'integrale de Fourier et ses applications a l'optique* (Besançon: private publication).

Hopkins, H. H., 1955, *Proc. Roy. Soc.* A, **231,** 91; 1956, *Proc. Phys. Soc.* B, **69,** 562; 1957, *Ibid.*, **70,** 449.

Lansraux, G., 1953, *Diffraction Instrumentale* (Paris : Editions de la Revue d'Optique), p. 69.

Linfoot, E. H., 1955, *Recent Advances in Optics* (Oxford : Clarendon Press.)

Miyamoto, K., 1957, *J. Appl. Phys.* (Japan), **26,** 421.

Steel, W. H., 1952, *Thesis*, University of Paris.

RESEARCH NOTES

A Physical Interpretation of Impedance for Rectangular Waveguides

By J. A. LANE
Radio Research Station, Slough, Bucks.

Communicated by R. L. Smith-Rose ; *MS. received* 22*nd August* 1957

In low-frequency electrical transmission, the characteristic impedance of transmission lines may be defined in terms of any two of the following quantities: voltage, current and transmitted power. For ordinary conditions all three definitions give identical numerical results. A similar procedure for waveguides, however, is generally regarded as having little practical significance since a different value of characteristic impedance results from each definition. For this reason, the concept of wave impedance η introduced by Schelkunoff (1943) is often used in calculating the reflection coefficient of waveguide discontinuities. η is defined as the ratio of the transverse components of electric and magnetic fields, and is constant over a given cross section for a given mode. Furthermore, if a transverse film having a surface resistivity equal to the wave impedance is followed by a reflecting plunger at a distance of one quarter of the guide wavelength, then the waveguide is terminated in a non-reflecting load. Many waveguide components, however, occupy only a small portion of the transverse cross section; for example, bolometer wires mounted parallel to the electric field of the dominant mode. For such components, it is of interest to examine the physical significance of the alternative definitions, i.e. those in terms of voltage, current and power. This note summarizes an experimental investigation of this aspect of waveguide impedance.

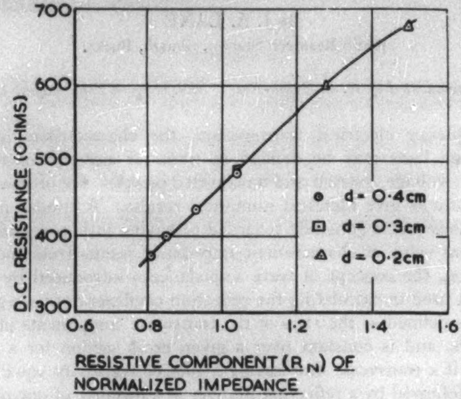
In the development of a film bolometer (Lane 1956) for microwave frequencies, it has been found that a non-reflecting waveguide load is provided by a thin strip of suitable resistivity located symmetrically in the transverse plane, and followed by a perfectly reflecting plunger. At a frequency of 9·2 Gc/s (wavelength 3·26 cm) the films used were of width 0·2–0·4 cm and consisted of a platinum deposit ($\sim 10^{-6}$ cm thick) on a thin mica support. Impedance measurements with such films show that the equivalent circuit consists of a resistance in series with a small inductance. For a film of given width, the resistivity can be reduced by careful heating of the film; the results of the impedance measurements can then be compared with the corresponding values of d.c. resistance.

Since the narrow films used are located in a region of maximum transverse voltage and are totally absorbing at an optimum value of resistivity, it is of particular interest to compare the results with a characteristic impedance, Z_{WV}, defined in terms of power and maximum transverse voltage. The figure shows the relation between the d.c. resistance of the film, as measured across the waveguide, and the resistive component of the measured normalized impedance, R_N. The results were obtained with films 0·2, 0·3 and 0·4 cm wide at a frequency of 9·2 Gc/s.

The value of d.c. resistance corresponding to $R=1$ is 470 ohms. The value of Z_{wv} is given by:

$$Z_{wv}=2\eta\lambda_g b/\lambda a \qquad \text{......(1)}$$

where λ_g = guide wavelength, λ = free-space wavelength, and a and b are the internal dimensions of the broad and narrow faces of the waveguide. For the conditions specified, $Z_{wv}=478$ ohms.



Resistance of transverse films for various widths d. $f=9{\cdot}2$ Gc/s.

The alternative values defined in terms of voltage and current (Z_{VI}), and current and power (Z_{WI}) are given by :

$$Z_{VI}=\eta\pi\lambda_g b/2a\lambda=376 \text{ ohms} \qquad \text{......(2)}$$

$$Z_{WI}=\eta\pi^2\lambda_g b/8a=295 \text{ ohms.} \qquad \text{......(3)}$$

These results indicate, therefore, that the most appropriate definition of waveguide impedance, for narrow resistive obstacles in rectangular waveguides at least, is one based on transmitted power and transverse voltage.

REFERENCES

LANE, J. A., 1956, *Nature, Lond.*, **177**, 392.

SCHELKUNOFF, S. A., 1943, *Electromagnetic Waves* (New York : Van Nostrand).

LETTERS TO THE EDITOR

Lifetimes of Excess Carriers in InSb

The recombination radiation from intrinsic indium antimonide has recently been investigated as follows. The crystal, at a temperature which is in excess of ambient, receives a primary beam of photons, which produces electron–hole pairs. Energy is radiated by the crystal at the rate at which it is received, and is interpreted as due to the recombination of pairs. The wavelengths of this emitted radiation can be filtered from the primary beam, and temperature effects can be removed by chopping. The numerical information obtained from this experiment has been used (Moss 1957, and references given there) to give an estimate $a=0{\cdot}21$ for the ratio of bulk to radiative lifetime τ_b/τ_r at room temperature.

From the optical absorption coefficient in the region of the absorption edge (about $7\,\mu$ for intrinsic material at room temperature), using the principle of detailed balance, the radiative lifetime can be estimated at $0{\cdot}7\,\mu$sec, and presumes a quantum efficiency of unity. The main contribution to the integration which is required to arrive at this value comes from the consideration of the absorption of photons whose energy is *less* than about 0·5 ev or about $3E$, where $E=0{\cdot}17$ ev is the main gap of indium antimonide (see, for instance, figure 1 of Mackintosh and Allen 1955). The bulk lifetime is about $0{\cdot}063\,\mu$sec (Goodwin 1956). This leads to a value $b=0{\cdot}09$ for the ratio of lifetimes at room temperature. Since this estimate is more direct than a, we shall regard it as more reliable. The estimate from the recombination radiation is then too large by a factor of about 2·3.

To account for this discrepancy, one may think of various possibilities: (i) Only a fraction γ of the non-equilibrium pairs recombines radiatively. (ii) Only a fraction β of the radiation actually received is due to the recombination of pairs. (iii) The quantum efficiency is not unity. The first effect is to be ruled out, since it would imply that the radiation received from the material every second is due to a smaller number of radiative recombinations than had been thought. This would make the inferred radiative lifetime smaller, and the estimate of the ratio of lifetimes even larger. As regards the second effect, it is difficult to see how a mechanism other than electron–hole recombination can give rise to a spectrum with a maximum near the wavelength corresponding to the main gap $E=0{\cdot}17$ ev, assuming temperature effects to have been effectively cycled out. The third possibility seems more likely. Assume that the radiation spectrum from the tungsten lamp, which was used in this experiment as the primary source, has a maximum for quanta of about 1 ev. This corresponds to an energy of $6E$. One knows from the work of Koc (1957) that the quantum efficiency of germanium at room temperature rises linearly from about 0·9 for quanta whose energy is $3{\cdot}2E'$ ($E'\sim 0{\cdot}72$ ev for germanium). At $6E'$ the quantum efficiency is already about 1·8. Thus it may not be unreasonable to suppose that the discrepancy between the two estimates of τ_b/τ_r is *in part* due to a mean quantum efficiency $\bar{\eta}$ which lies between 1 and 2 for indium antimonide at room temperature, when exposed to radiation from a tungsten lamp whose long wavelengths have been removed by glass lenses. This would mean that the rate at which radiation is emitted is due to a greater number of recombining pairs than had been thought. This makes the inferred radiative lifetime longer, and the estimated lifetime ratio smaller, as required.

It may be shown that, with the generalizations indicated above, one obtains the following result (which extends Moss's calculation) for recombination radiation data obtained with a strongly absorbed primary beam

$$\bar{\eta}\gamma\frac{\tau_b}{\tau_r}=\frac{\beta R_c}{FfI+\beta(1-f)R_c}.$$

R_c is the rate at which photons emitted from the material are collected at the receiver, I is the rate at which photons of the primary beam are absorbed, F is the fraction of the emitted photons which can be registered by the detecting apparatus, and f is the fraction of direct recombination photons (this excludes photons which have been produced by the absorption of recombination photons in the material) which can be emitted from the crystal without being internally absorbed. Moss gave $R_c=2{\cdot}6\times10^{13}$ sec^{-1}, $I=1{\cdot}3\times10^{17}$ sec^{-1}, $F=0{\cdot}004$, $f=0{\cdot}2$, and assumed $\bar{\eta}=\gamma=\beta=1$. With $1<\bar{\eta}\leqslant2$, one finds $0{\cdot}2>\tau_b/\tau_r\geqslant0{\cdot}1$.

This type of calculation is rather rough. Thus the composition of the primary beam enters only through its mean quantum efficiency $\bar{\eta}$ and through the steady non-equilibrium temperature T which the incident beam produces in the crystal. The extrapolation of Koc's result as ratios can also serve only as a guide.

Marischal College,
Aberdeen.
6th August 1957.

P. T. Landsberg.

Goodwin, D. W., 1956, *Report of the Rugby Meeting on Semiconductors* (London : Physical Society).
Koc, S., 1957, *Czech. J. Phys.*, **7**, 91.
Mackintosh, I. M., and Allen, J. W., 1955, *Proc. Phys. Soc.* B, **68**, 985.
Moss, T. S., 1957, *Proc. Phys. Soc.* B, **70**, 247.

Paramagnetic Resonance Absorption in Iodine

Paramagnetic resonance absorption has been observed in free atoms of ^{127}I, obtained by photo-dissociation of the molecular vapour within the cavity resonator. This method was suggested by Professor B. Bleaney. The apparatus used has already been described by Bowers, Kamper and Knight (1957). Six of the eighteen $\Delta M_I=0$, $\Delta M_J=1$ transitions of the $^2P_{3/2}$ ground state were measured with an accuracy of one or two parts per million.

It was possible to fit the results, after correction for second order effects, to the hyperfine structure interaction constants measured in zero field by Jaccarino, King, Satten and Stroke (1954), and a value of $438{\cdot}528\pm1\times10^{-3}$ for the ratio of the electronic g-factor to that of the proton. Taken with the result of Koenig, Prodell and Kusch (1952) this leads to a value of $0{\cdot}666225\pm1{\cdot}5\times10^{-6}$ for the ratio of the electronic g-factor to that of a free electron spin. An estimate of the relativistic and diamagnetic contributions to this quantity by the method of Abragam and Van Vleck (1953) has shown that there is no significant evidence for any breakdown of Russell–Saunders coupling. If this does occur, the admixtures are considerably less than one part in a thousand.

Using the intensity of the paramagnetic resonance spectrum as a measure of the concentration of free atoms, it was possible to investigate the dependence of this quantity on the conditions within the cavity. Two light sources were available; a mercury lamp and a xenon lamp, which could be used with appropriate filters to give light of wavelength respectively longer or shorter than the convergence limit of the visible band system of the iodine molecule at 4990 Å. Experiments were conducted below 1 mm of mercury pressure. It was found that if the partial pressure of molecular iodine was held constant at about 0·1 mm, and the total pressure increased by the addition of air, the concentration of atoms remained constant. It was also insensitive to changes in the partial pressure of molecular iodine in the range between 0·02 mm and 0·2 mm. These results are independent of whether dissociation is produced by light of wavelength shorter or longer than 4990 Å.

Since the rate of dissociation by light in the continuous part of the molecular spectrum is not affected by collisions, it follows that the rate of recombination is independent of the pressure. This being so, it also follows that the rate of dissociation by light in the banded region of the spectrum is approximately independent of the pressure. The results of Wood (1911) on the quenching of molecular iodine fluorescence show that any collision effect would be strongly pressure dependent in this range, so that these results indicate that a mechanism other than the collision induced predissociation exists by which light of wavelength greater than 4990 Å can dissociate the molecules, and that this mechanism becomes important at very low pressures. It was also found that a change by a factor of 2 in the magnetic field does not have any detectable effect on the intensity of the lines, so that magnetically induced predissociation (Turner 1930) is not making a significant contribution in this case.

The only recombination mechanism that appears to fit these results is a three-body collision in the gas phase with an iodine molecule acting as third body. This may be regarded as qualitative confirmation of the large rate constant for this process, as compared with one using air or argon as a third body, deduced by Christie, Harrison, Norrish and Porter (1955). Before recombination, the atoms survive of the order of ten collisions with the walls of the cell.

The collision diameters for line broadening by air and argon do not differ significantly from their kinetic theory values.

This work was undertaken as a preliminary to a study of the hyperfine structure of the ground state of ^{129}I.

Clarendon Laboratory,
Oxford.
7th October, 1957.

†K. D. Bowers.
R. A. Kamper.
C. D. Lustig.

†Now at Bell Telephone Laboratories, Murray Hill, N.J., U.S.A.

Abragam, A., and Van Vleck, J. H., 1953, *Phys. Rev.*, **92**, 1448.
Bowers, K. D., Kamper, R. A., and Knight, R. B. D., 1957, *J. Sci. Instrum.*, **34**, 49.
Christie, M. I., Harrison, A. J., Norrish, R. G. W., and Porter, G., 1955, *Proc. Roy. Soc.* A, **231**, 446.
Jaccarino, V., King, J. G., Satten, R. A., and Stroke, H. H., 1954, *Phys. Rev.*, **94**, 1798.
Koenig, S., Prodell, A. S., and Kusch, P., 1952, *Phys. Rev.*, **88**, 191.
Turner, L. A., 1930, *Z. Phys.*, **65**, 464.
Wood, R. W., 1911, *Phil. Mag.*, **21**, 309.

REVIEWS OF BOOKS

High Speed Photography, edited by R. B. Collins. Pp. xvi + 417. (London: Butterworths Scientific Publications, 1957.) 70*s*.

This volume contains the Proceedings of the Third International Congress on High Speed Photography, which was held under the auspices of the Department of Scientific and Industrial Research, from 10th to 15th September, 1956.

It is a pleasure to review a book that has been so ably edited and so well published. It includes a great deal of original research and information of a practical kind. It is particularly pleasing that the book has become available so shortly after the Congress.

In October 1952 an International Congress on High Speed Photography was held in Washington; a second was held in Paris in September 1954; the third Congress was held in London. Authors from eight different countries presented sixty four papers in all. These all describe techniques and applications that were original, and were not elsewhere available in published literature. There was at the Congress adequate time for the consideration of papers, and resumés of the vigorous discussion have been published along with each group of papers. The volume is extensively illustrated, and altogether forms a worthy successor to the excellent volume, *Actes du 2e Congrès International de Photographie et Cinématographie Ultra-rapides*, edited by P. Naslin and J. Vivie (Paris: Dunod, 1956). My only adverse criticism concerns the absence of an index.

The different chapters of the Proceedings are devoted to Flash Light Sources, Image Splitting and Image Sampling Techniques, Inertialess Shutters, Applications in Biology and Medicine and in Machine Analysis, Applications in Ballistics and Explosives, Instrument Aids, Photographic Materials, X-rays, Film Evaluation, Schlieren and Interferometric Techniques, Rotating Mirror Cameras, Medium Repetition Rate Cameras, Applications in Aerodynamics, Applications in Hydrodynamics, and Review Papers. The range of topics covered indicates the impressive growth and extent of the subject.

Photographic methods are amongst the most powerful laboratory aids. The study of even comparatively slow processes requires short exposures when our interest is in the fine detail, and increasingly it is necessary to make use of the techniques of high speed photography. In micrography the problem is accentuated and we must take advantage of the most advanced and elaborate of the methods. These in turn require continued research and development.

The publication of this volume makes available a wider acquaintance with the methods of high speed photography and of the ways these may be applied in industry and in fundamental research. There is throughout a most commendable emphasis on the way that techniques, which were developed for some particular study, may be extended and adapted for use in other fields.

The ever wider use of high speed photography in research and industry is a foregone conclusion. It is as certain too that this will contribute to more rapid progress in science and technology. J. S. COURTNEY-PRATT.

The Physics of Non-Destructive Testing: British Journal of Applied Physics Supplement No. 6. Pp. iii + 72. (London: Institute of Physics, 1957). 25*s.*

This volume appears as a Supplement to the British Journal of Applied Physics and contains a collection of papers presented at meetings of the Non-destructive Testing Group of the Institute of Physics. It is in the nature of a progress report on the Group's activities. While in no sense a comprehensive survey of the very broad subject of physics in non-destructive testing, it contains papers covering a wide diversity of topics and gives a useful picture of the fields in which effort is being directed by the Group.

The physicist or engineer concerned with non-destructive testing often faces a difficult and unenviable task. Broadly speaking he is required to extrapolate to $t \gg 0$ on the basis of non-destructive tests applied at $t \sim 0$. He must frequently do this for a system which is so complex that the underlying physical phenomena are only imperfectly understood—if they are understood at all. Many of the procedures which are established and of proven usefulness as non-destructive tests are of an empirical nature: the mechanism of operation is frequently obscure. It is to be expected that effort directed towards an understanding of the fundamentals behind such tests will in time bring dividends. At the same time, studies of the relations between other physical properties may reveal correlations which suggest new methods of testing. Both approaches are represented in the papers given in this volume. There are contributions with an academic bias and of a general nature discussing the relation between structures and other fundamental physical properties. Optical and magnetic properties are discussed in this way. The relatively new subject of nuclear magnetic resonance is shown to offer interesting possibilities as is also the subject of radiation damage in solids. X-ray diffraction techniques are shown to offer several promising lines of attack, especially those using high-voltage diffraction testing. A brief review of problems on fatigue is included, together with a report of work in which structure changes and energy dissipation are examined during fatigue of copper. The many problems associated with crack detection by the use of penetrants are well brought out in the record of a symposium on this subject. Attention is also directed to the testing of non-metals and fibres and—in a paper which is both entertaining and informative—on the testing (non-destructively) of humans. In this connection the advances in radiographic methods employing xeroradiography are shown in a striking manner.

Although the book is perhaps less comprehensive than the title suggests it should form an invaluable addition to the collection of the worker in this field and gives an excellent picture of the way in which the academic and the applied are being drawn together.

O. S. HEAVENS.

CONTENTS OF SECTION A

PAGE

Mr. A. M. Arthurs and Mr. J. Hyslop. Radiative Charge Transfer from H Atoms to He^{2+} Ions 849

Dr. D. G. Ravenhall and Dr. D. R. Yennie. Electron Scattering and Model Independence 857

Dr. D. L. Booth, Dr. R. S. Hill, Dr. F. V. Price and Dr. D. Roaf. The Reaction ^{3}He(d, p)^{4}He between 100 kev and 500 kev 863

Mr. S. J. Biel. The Binding Energy of ^{8}Be and ^{12}C 866

Dr. J. M. Anderson, Dr. A. D. Kavadas and Dr. R. W. McKay. The Decay of the Nitrogen Afterglow 877

Dr. J. M. Anderson. The Nature of Active Nitrogen 887

Research Notes :

Mr. S. Hinds, Dr. R. Middleton and Dr. G. Parry. Angular Distributions of Some Inelastically Scattered Deuteron Groups from ^{24}Mg and ^{28}Si . 900

Mr. S. Abid Husain and Mr. A. K. Hilmi. Absorption of Beta Particles in Liquids 902

Mr. W. Hayes, Mr. P. D. McCarvill and Mr. T. E. Nevin. The Structure of the $x^{7}\Sigma$ State of MnH and MnD 904

Dr. D. Park. An Improved Born Approximation 905

Letters to the Editor :

Dr. A. P. Stone. Expressions for certain Wigner Coefficients 908

Corrigendum :

Dr. R. D. Connor and Mr. I. L. Fairweather. The β-spectrum of Sulphur 35 909

Reviews of Books 910

Contents of Section B 912

Subject Index, Section A, Vol. 70 913

Index of Authors (with Titles), Section A, Vol. 70 919

Index to Reviews of Books, Section A, Vol. 70 927

Figure 1. Radio echoes from the moon, using a double transmitter pulse every 4·4 seconds. The spacing between the pulses could be switched between ¼, ½, ¾, 1, 1¼ and 1½ seconds, and in the above photograph was 1·0 seconds.

Plate I

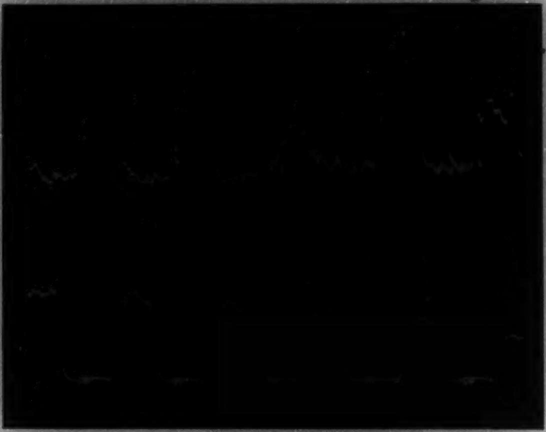


Figure 7 (*a*). A sample of echoes obtained using 10 msec transmitter pulses, with a receiver bandwidth of 500 c/s. The echoes are strobed from the time base and the negative going spikes are range marks. The variation in amplitude is due to rapid fading. Also shown are the transmitter pulses picked up direct by the receiver via a suitable attenuator. The variation in the shape of the top of the pulses is due to unwanted 100 c/s frequency modulation.

Figure 7 (*b*). A sample of 2 msec transmitter pulses and echoes. The echoes show no lengthening as would be expected if the moon behaved as a reflector whose depth were greater than about 1 msec.

PROCEEDINGS OF THE PHYSICAL SOCIETY
SECTION B, 1957—VOL. 70

SUBJECT INDEX

PAGE
Aberration permissible in optical systems 449
Absorption coefficient, microwave, temperature variation in ethyl chloride (R) 621
Absorption, *see also* Dielectric, Paramagnetic resonance
Absorption spectra of single crystals of pure and partially decomposed silver azide (R) 147
Absorption, ultra-violet, of counting diamonds, variations in 320
Adhesion theory of friction 98
Aerosol, concentration changes in 65
Air showers, extensive, time variation of 840
Alcohols, ethyl and methyl, dielectric properties in wavelength range 3 cm–52 cm 937
Alcomax, processes occurring during heat treatment of 817
Alcomax III, quenched, unusual magnetic properties of 823
Alloy-diffusion, process for making diffused-base junction transistors (R) 1087
Alloys, CuAl and CoNi, stacking faults and dislocations 1151
Alloys, Fe–Ni, complex permeability at high frequencies 186
Aluminium, anomalous form of sub-boundary in (R) 533
Analogue method for studying multiple scattering 605
Aplanat, two-mirror 945
Atomic heat of cerium between 1·5°K and 20°K 471
Atomic heat of samarium from 2°K to 20°K (R) 434

Bands of the $A^3\Pi_{0+}$–$X^1\Sigma^+$ systems of TlCl, rotational analysis (L) 622
Bitter figure technique, a new colloid for use with (R) 1000
Breakdown, electrical, of single crystals of KCl and NaCl, directional effects 169

Cadmium, single crystals, effect of thin films of gold on mechanical properties 608
Calcium, liquid, viscosity 1079
Carriers, excess, in InSb, lifetime (L) 1175
Carriers, free, in semiconductors, absorption of (R) 883
Carriers, minority, diffusion in presence of trapping 230
Carriers, minority, drift in presence of trapping 223
Cerium, atomic heat between 1·5°K and 20°K 471
Charge carriers in amorphous selenium films, transit time measurements of 669
Colloid, new, for use with Bitter figure technique (R) 1000
Compressional waves, high frequency, propagation in isotropic cylinders 136
Concentration changes in aerosol 65
Condenser aperture, influence on resolution of pair of parallel lines 361
Conduction currents, high field, in hexane 913
Conductivity junctions, low–high, in semiconductors (R) 885
Conductivity and dielectric constant of guanidine aluminium sulphate during switching 702
Contact effects from compression and flash sintering in graphite powders 928
Contact, electric, to zinc sulphide crystals (L) 1096
Contact, point, transistor studies with radioactive collectors (R) 145
Contact potential difference measurements by the Kelvin method 676
Crystal fields of high symmetry, splittings induced in levels by (R) 880
Crystals, cobalt, domain patterns on (R) 254
Crystals, nuclear relaxation mechanism in, method for determining (R) 436
Crystals, single, Cd, effect of thin films of gold on mechanical properties 608
Crystals, single, ferrite, domain patterns on (L) 441
Crystals, single, KCl and NaCl, directional effects in electrical breakdown 169
Crystals, single, nickel, saturation magnetostriction (L) 1093
Crystals, single, of pure and partially decomposed silver azide, absorption spectra (R) 147

PAGE

Crystals, single, *see also* Photodiode
Crystals, zinc sulphide, electric contact to (L) 1096
Crystals, *see also* Epitaxial growth
Cupric oxide, thermal decomposition *in vacuo* (R) 1005

Defects, quantum, extrapolated, as a check on calculating phases for scattering of electrons by positive ions (R) 620
Deposits, evaporated silver, structure when prepared at high temperatures (R) 251
Diamond, p-type semiconducting magnetoresistance (R) 527
Diamond surfaces, induction of ring cracks 521
Diamonds, counting, variations in ultra-violet absorption 320
Diamonds, type IIb, electrical and optical properties 177
Dielectric absorption, Maxwell–Wagner type, with wool wax–water mixtures (R) 427
Dielectric body, dipole moment fluctuations of 314
Dielectric constant and conductivity of guanidine aluminium sulphate during switching 702
Dielectric constant of mixtures, formulae 769
Dielectric properties of methyl and ethyl alcohols in the wavelength range 3 cm–52 cm 937
Diffusion, alloy, *see* Alloy-diffusion
Discharge, gas *see* Gas discharge
Discharges, high power, in intense longitudinal field 212
Dispersion effects in ultrasonic waveguides and importance in measurement of attenuation 721
Dislocations and stacking faults in CuAl and CoNi alloys 1151
Domain patterns on cobalt crystals (R) 254
Domain patterns on ferrite single cystals (L) 441
Domain sizes, ferromagnetic and surface structures 77

Eddy currents, effect on nuclear magnetic resonance in metals 345
Elastic moduli, dynamic, of organic glass, temperature variation 1045
Electrojet, equatorial effect on ionospheric E_s and F2 layers 833
Electron bombardment of thin barium films 235
Electron concentrations in plasma in presence of negative ions, measurement by r.f. probe technique 663
Electron diffraction, *see* Films
Electron microscopy of tin whiskers using carbon replicas (R) 889
Electron swarm in magnetic field, lateral diffusion 305
Electrons, interaction with lattice vibrations 390
Electrons, scattering by positive ions, the use of extrapolated quantum defects as a check on calculated phases of (R) 620
Energy levels, temperature dependent, in statistical mechanics 161
Epitaxial growth and oxidation of Ni, Co and Fe on rocksalt 265
Ethyl chloride, temperature variation of microwave absorption coefficient in (R) 621
Evaporated silver deposits prepared at high temperatures, structure (R) 251

Ferrimagnetism, and the Ising model (L) 537
Ferrites, effect of grain size on magnetothermal properties (R) 623
Ferromagnetic metals, magnetoresistance of Ni–Zn ferrite system (R) 531
Ferromagnetics, magnetothermal relations 1123
Ferromagnetism, surface structures and domain sizes 77
Field, longitudinal, intense, high power discharge in 212
Films, copper, vacuum deposited, stress annealing 950
Films, of Fe, Ni, Co on rocksalt, epitaxial growth and oxidation 265
Films, thin, barium, electron bombardment 235
Films, thin, of gold, effect on mechanical properties of Cd single crystals 608
Fission cross sections of ^{233}U, ^{235}U, ^{238}U and ^{239}Pu for neutrons in the energy range 0·03 Mev to 3.0 Mev 573
Fluid, ideally conducting hydromagnetic instabilities of 31

PAGE
Fourier images : I—Point source . . . 486
Fourier images : II—Out-of-focus patterns . . . 497
Fourier images : III—Finite sources . . . 505
Fractures, controlled, experiments on . . . 201
Frequency response, geometrical-optical treatment . . . 1162
Frequency response of optical systems, numerical evaluation (R) . . . 1002
Friction, adhesion theory . . . 98
Friction, surface, of solids, general theory (R) . . . 431
Friction, variation with velocity . . . 198
Fringes, white light, of equal chromatic order, formation (R) . . . 256
Functions induced in germanium surface by transverse electric fields . . . 481

Galactic radio sources, angular diameters at wavelength 10·7 cm (R) . . . 328
Gas, classical, of non-attracting rigid spheres, possibility of phase transition in (L) . 536
Gas bubbles, effect on sound propagation in water . . . 85
Gas consisting of hard spherical molecules, Clusius and Dickel constants . . . 93
Gas discharge, constricted, influence of axial magnetic field on stability . . . 1049
Gas discharge, constricted instability . . . 49
Gas discharge, *see also* Pinched discharge
Germanium, p-type, field dependence of mobility in (L) . . . 899
Germanium surfaces, junctions induced by transverse electric fields . . . 481
Glass, organic, temperature variation of dynamic elastic moduli at two frequencies. 1045
Gold manganide, Au$_2$Mn, resistivity and magnetoresistance (R) . . . 1089
Gold, self-diffusion in . . . 545
Guanidine aluminium sulphate, dielectric constant and conductivity during switching 702
Guanidine aluminium sulphate hexahydrate, polarization reversal process in 1064

Hall coefficient, *see under* Semiconductor
Harmonic forces and lattice vibrations in solids . . . 1143
Heat treatment of Alcomax, processes occurring during . . . 817
Helium isotopes, melting curves (R) . . . 150
Hexagonal close-packed lattice, high temperature susceptibility expansions for (L) . 896
Hexane, high field conduction currents in. . . . 913
Hydrogen isotopes, melting curves (R) . . . 150
Hydromagnetic instabilities of ideally conducting fluid . . . 31
Hyperfine structure and nuclear electric quadrupole moment of ^{17}O (L) . . . 897

Impedance, physical interpretation for rectangular waveguides (R) . . . 1173
Impurity centres, divalent, in semiconductors . . . corr. 447
InAs, photoelectric effects in, at room temperature (L) . . . 1011
Indium antimonide, current noise (L) . . . 331
Indium, liquid, viscosity . . . 1069
Inert elements, lattice constants of . . . 700
Inertial effects and relaxation processes : I—Free rotation about fixed axis . 402
Inertial effects and relaxation processes : II—Free rotation in space . 414
Infra-red luminescence of iodine (R) . . . 892
Iodine, infra-red luminescence of (R) . . . 892
Iodine, paramagnetic resonance absorption in (L) . . . 1176
Ionizing particles, detection by the neon flash tube . . . 687
Ionosphere, oblique ray paths in . . . 653
Ionosphere study, statistical distribution arising in . . . 559
Ionospheric E$_s$ and F2 layers, effect of equatorial electrojet on . . . 833
Ions, positive, collection by probe immersed in plasma . . . 297
Iron ditelluride, thermal expansion . . . 1113
Ising model and ferrimagnetism (L) . . . 537

Jets, high speed metallic, penetration by . . . 867

PAGE

Kelvin method for the measurement of contact potential differences . . . 676

Lattice constants of inert elements . . . 700
Lattice plane square, Onsager–Ising partition function, influence of boundary conditions . . . 192
Lattice vibrations and harmonic forces in solids . . . 1143
Lattice vibrations, interaction of electrons with . . . 390
Lead selenide, reaction of oxygen with . . . 1025
Light distribution in image of two parallel lines, influence of condenser aperture . . . 361
Light, extinction by suspensions of silica dust . . . 102
Light scattering by suspensions of randomly oriented long prisms, theory . . . 379
Light, white, formation of fringes of equal chromatic order (R) . . . 256
Lightning, inter-stroke processes, radar echoes from . . . 961
Luminescence, infra-red, of iodine (R) . . . 892

Magnesium, liquid, viscosity . . . 1079
Magnetic field, lateral diffusion of electron swarm in . . . 305
Magnetic fields, strong, of short duration, production and intensity measurement (R) 432
Magnetic properties, unusual, of quenched Alcomax III . . . 823
Magnetic refrigerator for temperatures below 1°K . . . 641
Magnetic susceptibilities of La, Ce, Pr, Nd and Sm from $1{\cdot}5^\circ$K to 300°K . . . 566
Magnetic susceptibility of ytterbium from $1{\cdot}3^\circ$K to 300°K . . . 476
Magnetoresistance of Ni–Zn ferrite system (R) . . . 531
Magnetoresistance of p-type semiconducting diamond (R) . . . 527
Magnetoresistance and resistivity of Au_2 Mn (R) . . . 1089
Magnetoresistance in strong magnetic fields (R) . . . 1008
Magnetostriction, saturated, of nickel single crystals (L) . . . 1093
Magnetothermal effects in Ni and Co . . . 51
Magnetothermal properties of ferrites, effect of grain size (R) . . . 623
Magnetothermal relations for ferromagnetics . . . 1123
Manuel model for MnBi, comments . . . 711
Melting curves of helium and hydrogen isotopes (R) . . . 150
Metal fatigue at high frequency . . . 393
Metal rectifiers, *see under* Rectifiers . . . 991
Metallic wear, *see* Wear, metallic
Metals, nuclear magnetic resonance in, effect of eddy current on . . . 345
Microwave absorption coefficient, temperature variation in ethyl chloride (R) . . . 621
MnBi, comments on Manuel model of . . . 711
Mobility, field dependence of in p-type germanium (L) . . . 899
Mobility in semiconductors, field dependence (L) . . . 628
Molybdenum, new phase structure of . . . 708
Moon, scattering of radio waves by . . . 1105

Neon flash tube as a detector of ionizing particles . . . 687
Nickel, temperature coefficient of reflectivity of . . . 862
Nuclear electric quadrupole moment and hyperfine structure of ^{17}O (L) . . . 897
Nuclear magnetic resonance in Eu^{3+}, theory . . . 119
Niclear magnetic resonance in metals, effect of eddy currents . . . 345
Nuclear magnetic resonance spectrum for isosceles triangular configurations of nuclei 980
Nuclear relaxation mechanism in crystals, method for determining (R) . . . 436

Onsager–Ising partition function for plane square lattice, influence of boundary conditions . . . 192

Optical properties of zinc sulphide, measurement of . . . 845
Optical systems, aberration permissible . . . 449
Optical systems, numerical evaluation of frequency response (R) . . . 1002
Optics, geometrical-optical treatment of frequency response . . . 1162
Overhauser effect in solids . . . 785

PAGE
Oxygen effect on an evaporated PbSe layer 704
Oxygen, reaction with lead selenide 1025

Paramagnetic resonance absorption in $GdCl_3$ 514
Paramagnetic resonance absorption in iodine (L) 1176
Paramagnetic resonance of impurities in NaF (L) 626
PbSe layer, evaporated, effect of oxygen on 704
Permeabilities, complex, of Fe–Ni alloys at high frequencies 186
Perspex, dynamic stress–optic coefficient 1033
Phase changes of second order in superconductors 553
Phase structure, new, of molybdenum 708
Phase transition in classical gas of non-attracting rigid spheres (L) 536
Phosphors, organic, photo-fluorescence decay times (R) 144
Photodiode, single crystal, of lead sulphide (L) 258
Photoelectric effects in InAs at room temperature (L) 1011
Photofluorescence decay times of organic phosphors (R) 144
Pinched discharge cooled by bremsstrahlung radiation 11
Pinched discharge, transient elementary theory 24
Pinched discharge, *see also under* Gas discharge
Plasma, collection of positive ions by immersed probe 297
Plasma electron concentrations, r.f. probe technique for measurement in presence of negative ions 663
Point contact, *see* Contact
Polarization reversal process in guanidine aluminium sulphate hexahydrate 1064
Powders, graphite, contact effects from flash sintering and compression 928
Probe technique, r.f., for measurement of plasma electron concentrations in presence of negative ions 663

Quadripole moment, nuclear electric, and hyperfine structure of ^{17}O (L) 897

Radar echoes from inter-stroke processes in lightning 961
Radiation, recombination, from InSb, theory of spectral distribution (R) 247
Radiative capture of alpha particles in ^{15}N 661
Radio sources, galactic, angular diameters at wavelength 10·7 cm (R) 328
Radio waves, scattering by the moon 1105
Rare earths, various, magnetic susceptibilities from 1·5°K to 300°K 566
Rare gas mixtures, binary and ternary, thermal conductivity 369
Rectifiers, metal, current–capacitance characteristics 991
Reflectivity, of nickel, temperature coefficient of 862
Relaxation processes and inertial effects: I—Free rotation about fixed axis 402
Relaxation processes and inertial effects: II—Free rotation in space 414
Relaxation, ultrasonic, in methylene chloride 71
Resistivity and magnetoresistance of Au_2Mn (R) 1089
Resonance, nuclear magnetic, *see* Nuclear magnetic resonance
Resonance, paramagnetic, *see* Paramagnetic resonance 980
Ring cracks on diamond surfaces, induction of 521

Scattering, light, *see under* Light
Scattering of radio waves by the moon 1105
Selenium films, amorphous, transit time measurements of charge carriers in selenium films 669
Self-diffusion in gold 545
Semiconducting diamond, p-type magnetoresistance (R) 527
Semiconductor, mixed, dependence of Hall coefficient on magnetic induction (InSb) 124
Semiconductors, the absorption of free carriers in (R) 883
Semiconductors, contributions to theory of electrical conductivity, thermal conductivity and thermoelectric power 590
Semiconductors, defects with several trapping levels corr. 154

PAGE

Semiconductors, excess carriers, recombination statistics 282
Semiconductors, field dependence of mobility in (L) 628
Semiconductors, fluctuations in number of electrons and holes (R) 244
Semiconductors, galvanomagnetic and thermomagnetic effects in (R) . . . 240
Semiconductors, InSb, spectral distributions of recombination radiation (R) . . 247
Semiconductors, low–high conductivity junctions in (R) 885
Semiconductors, statistics of divalent impurity centres. . . . . . corr. 447
Semiconductors, *see also* Carriers, Germanium, Indium antimonide, Selenium
Showers, extensive air, time variation of 840
Single crystals, *see* Crystals, single
Sintering, flash and compression in graphite powders, contact effects . . . 928
Sodium fluoride, paramagnetic resonance of impurities in (L) 626
Solid solution in GaAs–InAs system (L) 153
Solids, lattice vibrations and harmonic forces 1143
Solids, overhauser effect in 785
Solids, surface friction, general theory (R). 431
Solids, thermal expansion 827
Sound propagation in water, effect of gas bubbles 85
Spectra, absorption, *see* Absorption
Splittings induced in levels by crystal fields of high symmetry (R) 880
Statistical mechanics of the steady state 577
Statistical mechanics, temperature dependent energy levels 161
Steady state, statistical mechanics of. 577
Stress annealing in vacuum deposited copper films 950
Stress, dynamic stress–optic coefficient of Perspex 1033
Superconductors, phase changes of second order. 553
Surface structures and ferromagnetic domain sizes 77
Surface tension of liquid tungsten by drop-weight method 1040
Susceptibilities, high temperature (R) 326
Susceptibility, high temperature expansions for close-packed hexagonal lattice (L) . 896
Susceptibility, magnetic, *see* Magnetic susceptibility
Suspensions of randomly oriented long prisms, theory of light scattering by . . 379
Suspensions of silica dust, extinction of light by. 102
System, GaAs–InAs, solid solution in (L) 153

Temperature, influence on metallic wear. 112
Thermal conductivity of binary and ternary rare gas mixtures 369
Thermal decomposition of cupric oxide *in vacuo* (R) 1005
Thermal diffusion, Clusius and Dickel column constants for gas consisting of hard spherical molecules 93
Thermal expansion of iron ditelluride 1113
Thermal expansion of solids 827
Thermonuclear reaction rates. 1
Thermonuclear reactor, power producing, criteria for 6
Thin foils of Al, anomalous form of sub-boundary in (R) 533
Tin, grey tin⇌white tin transformation in commercially pure tin, effect of liquid media 1135
Transistor, point contact, studies with radioactive collecters (R) 145
Transistors, diffused-base, process for making (R) 1087
Transistors, junction, abnormal noise in during secondary ionization (R). . . . 251
Tritons and α-particles emitted from deuteron bombardment of teflon magnetic analysis 233
Tungsten, liquid, surface tension, determination by drop-weight method . . . 1040

Ultrasonic relaxation in methylene chloride 71

Viscosity of liquid helium and liquid tin 1069
Viscosity of liquid magnesium and liquid calcium 1079

PAGE
Waveguides, rectangular, impedance, physical interpretation (R) 1173
Waveguides, ultrasonic, dispersion effects and importance in measurement of attenuation 721
Wear, metallic, influence of temperature 112

Ytterbium, magnetic susceptibility from 1·3 to 300°K 476

Zinc sulphide, crystals, electric contact to (L) 1096
Zinc sulphide, measurement of the optical properties of 845

INDEX OF AUTHORS (WITH TITLES)

PAGE

Aggarwal, P. S., and Goswami, A. : A new phase structure of molybdenum (R) 708

Alfrey, G. F., and Cooke, I. : Electric contact to zinc sulphide crystals (R) . . 1096

Allen, J. E. : An elementary theory of the transient pinched discharge . . . 24

Allen, J. E., Boyd, R. L. F., Reynolds, P. : The collection of positive ions by a probe immersed in a plasma 297

Anderson, J. C., and Donovan, B. : On the complex permeability of iron–nickel alloys at high frequencies 186

Andreae, J. H. : Ultrasonic relaxation in methylene chloride 71

Andrew, E. R., and Finch, N. D. : Nuclear magnetic resonance spectrum for isosceles triangular configurations of nuclei 980

Andrew, E. R., and Swanson, K. M. : A method for determining the nuclear relaxation mechanism in crystals (R) 436

Apostolakis, A. J., and Macpherson, I. : An emulsion spectrograph—a new technique for the determination of the momenta of fast charged particles : I—Design consideration and theory 146

Apostolakis, A. J., and Macpherson, I. : An emulsion spectrograph—a new technique for the determination of the momenta of fast charged particles : II—Construction and applications of a normal incidence stack 154

Banbury, P. C., with Nixon, J. D. : Junctions induced in germanium surfaces by transverse electric fields 481

Barakat, N. : The formation of white light fringes of equal chromatic order (R) . 256

Barrie, R. : Magnetoresistance in strong magnetic fields (R) 1008

Bate, G., with Starkiewicz, J., Bennett, H., and Hilsum, C. : A single crystal photo-diode of lead sulphide (L) 258

Batty, C. J., and Goldsack, S. J. : Polarization of protons scattered from carbon at 970 MEV (L) 165

Beale, J. R. A. : Alloy-diffusion : A process for making diffused-base junction transistors (R) 1087

Benbow, J. J., and Roesler, F. C. : Experiments on controlled fractures . . . 201

Bennett, H., with Starkiewicz, J., Bate, G., and Hilsum, C. : A single crystal photo-diode of lead sulphide 258

Bickerton, R. J. : The lateral diffusion of an electron swarm in a magnetic field . 305

Birss, R. R. : The saturation magnetostriction of nickel single crystals (R) . . 1093

Blackman, M. : On the thermal expansion of solids 827

Bleaney, B., and Hayes, W. : Paramagnetic resonance of impurities in NaF (L) . 626

Bogle, G. S., Dodd, J. N., and McLean, W. L. : Hyperfine structure coupling constant of the 6s6p^{3}P$_1$ level of ^{199}Hg by double resonance 796

Bowers, K. D., Kamper, R. A., and Lustig, C. D. : Paramagnetic resonance absorption in iodine (L) 1176

Boyd, R. L. F., with Allen, J. E., and Reynolds, P. : The collection of positive ions by a probe immersed in a plasma 297

Bradley, D. E., Franks, J., and Rush, P. E. : Electron microscopy of tin whiskers using carbon replicas (R) 889

Bradley, E. D., and Phillips, R. : An explanation of an anomalous form of sub-boundary in aluminium (R) 533

Bray, R., with Mendelson, K. S. : Field dependence of mobility in p-type germanium (L) 899

Buckingham, A. D. : The influence of a strong magnetic field on the dielectric constant of a diamagnetic fluid 753

Bullock, F. W. : Nuclear disintegrations produced by cosmic-ray particles in a high pressure cloud chamber 134

PAGE

Calverley, A. : A determination of the surface tension of liquid tungsten by the drop-weight method 1040
Carruthers, R., and Davenport, P. A. : Observations of the instability of constricted gaseous discharges 49
Champion, F. C., and Humphreys, D. L. O. : Variations in the ultra-violet absorption of counting diamonds 320
Champion, K. S. W. : High power discharges in an intense longitudinal field 212
Chapman, A. C., Rhodes, P., and Seymour, E. F. W. : The effect of eddy currents on nuclear magnetic resonance in metals 345
Cheesman, G. H., and Soane, C. M. : The lattice constants of the inert elements (R) 700
Christian, J. W., and Spreadborough, J. : Stacking faults and dislocations in copper–aluminium and cobalt–nickel alloys 1151
Christoffel, D. A. : The effect of grain size on the magnetothermal properties of ferrites (R) 623
Clegg, A. G., and McCaig, M. : Processes occurring during the heat treatment of Alcomax 817
Collins, L. E., and Heavens, O. S. : The epitaxial growth and oxidation of nickel, cobalt and iron on rocksalt 265
Coogan, C. K. : The measurement of the optical properties of zinc sulphide 845
Cooke, I., with Alfrey, G. F. : Electric contact to zinc sulphide crystals (R) . 1096
Cooper, R., Grossart, D. T., and Wallace, A. A. : Directional effects in the electric breakdown of single crystals of KCl and NaCl 169
Cowley, J. M., and Moodie, A. F. : Fourier images : I—The point source . . 486
Cowley, J. M., and Moodie, A. F. : Fourier images : II—The out-of-focus patterns 497
Cowley, J. M., and Moodie, A. F. : Fourier images : III—Finite sources . . 505
Craik, D. J., and Griffiths, P. M. : A new colloid for use with the Bitter figure technique (R) 1000
Crawford, S. M. : The variation with temperature of the dynamic elastic moduli of an organic glass at two frequencies 1045
Culpin, M. F. : The visocosity of liquid indium and liquid tin 1069
Culpin, M. F. : The viscosity of liquid magnesium and liquid calcium . . 1079

Danielian, A., and Stevens, K. W. H. : High temperature susceptibilities (R) . . 326
Daunt, J. G. : The magnetic refrigerator for temperatures below 1°K . . 641
Davenport, P. A., with Carruthers, R. : Observations of the instability of constricted gaseous discharges 49
Davies, L. W. : Low-high conductivity junctions in semiconductors (R) . . 885
Dodd, J. N., with Bogle, G. S., and McLean, W. L. : Hyperfine structure coupling constant of the $6s6p\,^3P_1$ level of ^{199}Hg by double resonance . . 796
Domb, C. : Melting curves of helium and hydrogen isotopes (R) 150
Domb, C., and Sykes, M. F. : High temperature susceptibility expansions for the close packed heaxgonal lattice (L) 896
Donovan, B., and March, N. H. : On the absorption by free carriers in semiconductors (R) 883
Donovan B., with Anderson, J. C. : On the complex permeability of iron–nickel alloys at high frequencies 186
Dryden, J. S., and Meakins, R. J. : Examples of the Maxwell–Wagner type of dielectric absorption using wool wax–water mixtures (R) 427

Elcock, E. W., and Landsberg, P. T. : Temperature dependent energy levels in statistical mechanics 161
Elliott, R. J. : Theory of nuclear magnetic resonance in Eu^{3+} 119
Ellison, J., McK. : Extinction of light by suspensions of silica dust . . . 102
Evans, J. V. : The scattering of radio waves by the moon 1105

Farley, F. J. M., and Storey, J. R. : Time variations of extensive air showers . . 840
Finch, N. D., with Andrew, E. R. : Nuclear magnetic resonance spectrum for isosceles triangular configurations of nuclei 980

PAGE

Foreman, A. J. E., and Lomer, W. M.: Lattice vibrations and harmonic forces in solids 1143
Franks, J.: The infra-red luminescence of iodine (R) 892
Franks, J., with Bradley, D. E., and Rush, P. E.: Electron microscopy of tin whiskers using carbon replicas (R) 889

Gardener, M., Kisdnasamy, S., Rössle, E., and Wolfendale, A. W.: The neon flash tube as a detector of ionizing particles 687
Gayther, D. B., and Nicholson, K. P.: The average neutron total cross sections of heavy elements at low energy 51
Gibalewicz, J., with Piekara, A., Malecki, J., and Surma, M.: Note on production of strong magnetic fields of short duration and measurement of their intensity (R) 432
Goldsack, S. J., with Batty, C. J.: Polarization of protons scattered from carbon at 960 Mev (L) 165
Goodbody, A. M.: The influence of condenser aperture on the resolution of a pair of parallel lines 361
Goswami, A., and Trehan, Y. N.: The thermal decomposition of cupric oxide *in vacuo* (R) 1005
Goswami, A., with Aggarwal, P. S.: A new phase structure of molybdenum (R) 708
Grant, E. H.: The dielectric properties of methyl and ethyl alcohols in the wavelength range 3 cm–52 cm 937
Griffiths, P. M., with Craik, D. J.: A new colloid for use with the Bitter figure technique (R) 1000
Grimley, T. B.: Homopolar effects in crystals having the sodium chloride type of lattice 123
Grossart, D. T. with Copper, R., and Wallace, A. A.: Directional effect in the electric breakdown of single crystals of KCl and NaCl 169

Hall, E. O.: Domain patterns on cobalt crystals (R) 254
Hamilton, T. D. S.: Photofluorescence decay times of organic phosphors (R) 144
Haneman, D., and Mortlock, A. J.: Point contact transistor studies using radioactive collectors (R) 145
Harper, P. G.: The interaction of electrons with lattice vibrations 390
Haselgrove, J.: Oblique ray paths in the ionosphere 653
Hawkins, T. D. F., with Moss, T. S., and Smith, S. D.: Absorption and dispersion of indium antimonide 776
Hayes, W., with Bleaney, B.: Paramagnetic resonance of impurities in NaF (L) 626
Head, A. K.: The two-mirror aplanat 945
Heavens, O. S., with Collins, L. E.: The epitaxial growth and oxidation of nickel cobalt and iron on rocksalt 265
Heikes, R. R., and Roberts, B. W.: Comments on the Manuel model for MnBi (L) 711
Herbstein, F. H.: A note on the structure of evaporated silver deposits prepared at high temperatures (R) 251
Hilsum, C.: Photoelectric effects in InAs at room temperature (L) 1011
Hilsum, C., with Starkiewicz, J., Bate, G., and Bennett, H.: A single crystal photodiode of lead sulphide 258
Hoffman, R. W., with Story, H. S.: Stress annealing in vacuum deposited copper films 950
Hopkins, H. H.: The aberration permissible in optical systems 449
Hopkins, H. H.: Geometrical optical treatment of frequency response 1162
Hopkins, H. H.: The numerical evaluation of the frequency response of optical systems (R) 1002
Hough, J. M., with Reynolds, J. A.: Formulae for dielectric constant of mixtures 769
House, H.: High field conduction currents in hexane 913
Howarth, D. J., Jones, R. H., and Putley, E. H.: The dependence of the Hall coefficient of a mixed semiconductor upon magnetic induction as exemplified by indium antimonide 124
Howes, V. R., with Tolansky, S.: Induction of ring cracks on diamond surfaces 521
Humphreys, D. L. O., with Champion, F. C.: Variations in the ultraviolet absorption of counting diamonds 320

PAGE

Hurst, C. : Phase changes of the second order 553
Hutchison, C. A., Judd, B. R., and Pope, D. F. D. : Paramagnetic resonance absorption in gadolinium trichloride 514

Jackson, R. B., and Walton, A. K. : Abnormal noise in junction transistors during secondary ionization (R) 251
Jacob, L. : The electron bombardment of thin barium films 235
Jones, R. H. : The effect of oxygen on an evaporated PbSe layer (R) . . . 704
Jones, R. H. : The reaction of oxygen with lead selenide 1025
Jones, R. H., with Howarth, D. J., and Putley, E. H. : The dependence of the Hall coefficient of a mixed semiconductor upon magnetic induction as exemplified by indium antimonide 124
Jonscher, A. K. : Diffusion of minority carriers in the presence of trapping . . 230
Jonscher, A. K. : Drift of minority carriers in the presence of trapping . . . 223
Judd, B. R. : The splittings induced in levels by crystal fields of high symmetry (R) . 880
Judd, B. R., with Hutchison, C. A., and Pope, D. F. D. : Paramagnetic resonance absorption in gadolinium trichloride 514

Kamper, R. A., Lea, K. R., and Lustig, C. D. : Hyperfine structure and nuclear electric quadrupole moment of ^{17}O (L) 897
Kamper, R. A., with Bowers, K. D., and Lustig, C. D. : Paramagnetic resonance absorption in iodine (L) 1176
King, R., with Lipsett, F. R. : The effect of thin films of gold on the mechanical properties of cadmium single crystals 608
Kisdnasamy, S., with Gardener, M., Rössle, E., and Wolfendale, A. W. : The neon flash tube as a detector of ionizing particles 687
Krishnaji, and Srivastava, G. P. : Temperature variation of microwave absorption coefficient in ethyl chloride (R) 621

Lamb, J., with Redwood, M. : On the propagation of high frequency compressional waves in isotropic cylinders 136
Lancaster, J. K. : The influence of temperature on metallic wear 112
Landsberg, P. T. : A contribution to the recombination statistics of excess carriers in semiconductors 282
Landsberg, P. T. : Defects with several trapping levels in semiconductors . corr. 154
Landsberg, P. T. : Lifetimes of excess carriers in InSb (L) 1175
Landsberg, P. T., with Elcock, E. W. : Temperature dependent energy levels in statistical mechanics 161
Lane, J. A. : A physical interpretation of impedance of rectangular waveguides (R) . 1173
Lawson, J. D. : Some criteria for a power producing thermonuclear reactor . . 6
Lea, K. R., with Kamper, R. A., and Lustig, C. D. : Hyperfine structure and nuclear electric quadrupole moment of ^{17}O (L) 897
LeClaire, A. D., with Makin, S. M., and Rowe, A. H. : Self-diffusion in gold . . 545
Lewis, F. A., Orr, J., and Ubbelohde, A. R. : Contact effects resulting from compression and flash sintering in graphite powders 928
Lipsett, F. R., and King, R. : The effect of thin films of gold on the mechanical properties of cadmium single crystals 608
Little, W. A. : The Overhauser effect in solids 785
Llewellyn, J. P., and Smith, T. : The thermal expansion of iron ditelluride . . 1113
Lock, J. M. : The magnetic susceptibilities of lanthanum, cerium, praseodymium, neodymium and samarium from 1·5°K to 300°K 566
Lock, J. M. : The magnetic susceptibility of ytterbium from 1·3°K to 300°K . . 476
Lomer, W. M., with Foreman, A. J. C. : Lattice vibrations and harmonic forces in solids 1143
Longuet-Higgins, M. S. : A statistical distribution arising in the study of the ionosphere 559
Lustig, C. D., with Bowers, K. D., and Kamper, R. A. : Paramagnetic resonance absorption in iodine (L) 1176

PAGE

Lustig, C. D., with Kamper, R. A., and Lea, K. R.: Hyperfine structure and nuclear electric quadrupole moment of ^{17}O (L) . . . 897

McCaig, M.: The unusual magnetic properties of quenched Alcomax III . . . 823
McCaig, M., with Clegg, A. G.: Processes occurring during the heat treatment of Alcomax . . . 817
McLaren, A. C.: The absorption spectra of single crystals of pure and partially decomposed silver azide (R) . . . 147
McLean, W. L., with Bogle, G. S., and Dodd, J. N.: Hyperfine structure coupling constant of the 6s6p^{3}P$_1$ level of ^{199}Hg by double resonance . . . 796
Macpherson, I., with Apostolakis, A. J.: An emulsion spectrograph—a new technique for the determination of the momenta of fast charged particles. Part I: Design consideration and theory . . . 146
Macpherson, I., with Apostolakis, A. J.: An emulsion spectrograph—a new technique for the determination of the momenta of fast charged particles. Part II: Construction and applications of a normal incidence stack . . . 154
Macpherson, J. D.: The effect of gas bubbles on sound propagation in water . . . 85
Makin, S. M., Rowe, A. H., and LeClaire, A. D.: Self-diffusion in gold . . . 545
Malecki, J., with Piekara, A., Surma, M., and Gibalewicz, J.: Note on production of strong magnetic fields of short duration and measurement of their intensity (R) . . . 432
Mansfield, R.: The galvanomagnetic and thermomagnetic effects in semiconductors (R) . . . 240
March, N. H., with Donovan, B.: On the absorption by free carriers in semiconductors (R) . . . 883
Marshall, D. F.: The dynamic stress–optic coefficient of Perspex . . . 1033
Martin, D. H.: Surface structures and ferromagnetic domain sizes . . . 77
Meakins, R. J., with Dryden, J. S.: Examples of the Maxwell–Wagner type of dielectric absorption using wool wax–water mixtures (R) . . . 427
Mendelson, K. S., and Bray, R.: Field dependence of mobility in p-type germanium (L) . . . 899
Mitchell, E. W. J., and Wedepohl, P. T.: Magnetoresistance of a p-type semiconducting diamond (R) . . . 527
Moriguchi, Yoshiro, and Okazaki, Atuyosi: Current–capacitance characteristics of metal rectifiers . . . 991
Mortlock, A. J., with Haneman, D.: Point contact transistor studies using radioactive collectors (R) . . . 145
Moodie, A. F., with Cowley, J. M.: Fourier images: I—The point source . . . 486
Moodie, A. F., with Cowley, J. M.: Fourier images: II—The out-of-focus patterns . . . 497
Moodie, A. F., with Cowley, J. M.: Fourier images: III—Finite sources . . . 505
Moss, T. S.: Theory of the spectral distribution of recombination radiation from InSb (R) . . . 247
Moss, T. S., Smith, S. D., and Hawkins, T. D. F.: Absorption and dispersion of indium antimonide . . . 776

Neppiras, E. A.: Metal fatigue at high frequency . . . 393
Newcomb, T. P., with Spurr, R. T.: The adhesion theory of friction . . . 98
Newcomb, T. P., with Spurr, R. T.: The variation of friction with velocity . . . 198
Nixon, J. D., and Banbury, P. C.: Junctions induced in germanium surfaces by transverse electric fields . . . 481

O'Dwyer, J. J.: The high temperature dielectric breakdown of alkali halides . . . 761
Okazaki, Atuyosi, with Moriguchi, Yoshiro: Current–capacitance characteristics of metal rectifiers . . . 991
Oliver, D. J.: Current noise in indium antimonide (L) . . . 331
Oliver, D. J.: Fluctuations in the number of electrons and holes in a semiconductor (R) . . . 244
Orr J., with Lewis, F. A., and Ubbelohde, A. R.: Contact effects resulting from compression and flash sintering in graphite powders . . . 928

PAGE

Paranjape, B. V.: Field dependence of mobility in semiconductors (L) . . . 628
Parker, R.: The magnetoresistance of the nickel–zinc ferrite system (R) . . . 531
Parkinson, D. H., and Roberts, L. M.: The atomic heat of cerium between 1·5°K and 20°K . . . 471
Parrott, J. E.: Some contributions to the theory of electric conductivity, thermal conductivity and thermoelectric power in semiconductors . . . 590
Pearson, R. F.: Domain patterns on ferrite single crystals (L) . . . 441
Pease, R. S.: Equilibrium characteristics of a pinched gas discharge cooled by bremsstrahlung radiation . . . 11
Phillips, R., with Bradley, D. E.: An explanation of an anomalous form of sub-boundary in aluminium (R) . . . 533
Piekara, A., Malecki, J., Surma, M., and Gibalewicz, J.: Note on production of strong magnetic fields of short duration and measurement of their intensity (R) . . . 432
Pope, D. F. D., with Hutchison, C. A., and Judd, B. R.: Paramagnetic resonance absorption in gadolinium trichloride . . . 514
Prutton, M.: Dielectric constant and conductivity of guanidine aluminium sulphate during switching (R) . . . 702
Prutton, M.: The polarization reversal process in guanidine aluminium sulphate hexahydrate . . . 1064
Putley, E. H., with Howarth, D. J., and Jones, R. H.: The dependence of the Hall coefficient of a mixed semiconductor upon magnetic induction as exemplified by indium antimonide . . . 124

Raynor, G. V., with Smith, R. W.: The effect of liquid media on the grey tin⇌white tin transformation in commercially pure tin . . . 1135
Redwood, M.: Dispersion effects in ultrasonic waveguides and their importance in the measurement of attenuation . . . 721
Redwood, M., and Lamb, J.: On the propagation of high frequency compressional waves in isotropic cylinders . . . 136
Reynolds, J. A., and Hough, J. M.: Formulae for dielectric constant of mixtures . . . 769
Reynolds, P., with Allen, J. E., and Boyd, R. L. F.: The collection of positive ions by a probe immersed in a plasma . . . 297
Rhodes, P., with, Chapman, A. C., and Seymour, E. F. W.: The effect of eddy currents on nuclear magnetic resonance in metals . . . 345
Riviere, J. C.: Contact potential difference measurements by the Kelvin method . . . 676
Roberts, B. W., with Heikes, R. R.: Comments on the Manuel model for MnBi (L) . . . 711
Roberts, L. M.: The atomic heats of calcium, strontium and barium between 1·5°K and 20°K . . . 738
Roberts, L. M..: The atomic heats of lithium, sodium and potassium between 1·5°K and 20°K . . . 744
Roberts, L. M.: The atomic heat of samarium from 2°K to 20°K (R) . . . 434
Roberts, L. M., with Parkinson, D. H.: The atomic heat of cerium between 1·5°K and 20°K . . . 471
Roesler, F. C., with Benbow, J. J.: Experiments on controlled fractures . . . 201
Rose, F. W. G.: The mass action laws for the reactions between free carriers and impurities in semiconductors considering the electron spin (R) . . . 801
Rössle, E., with Gardener, M., Kisdnasamy, S., and Wolfendale, A. W.: The neon flash tube as a detector of ionizing particles . . . 687
Rowe, A. H., with Makin, S. M., and LeClaire, A. D.: Self-diffusion in gold . . . 545
Rowlands, G., with Teale, R. W.: Some magnetothermal relations for ferromagnetics 1123
Rowson, B.: Measurements of the angular diameters of two galactic radio sources at a wavelength of 10·7 cm (R) . . . 328
Rush, P. E., with Bradley, D. E., and Franks, J.: Electron microscopy of tin whiskers using carbon replicas (R) . . . 889

Sack, R. A.: Relaxation processes and inertial effects: I—Free rotation about a fixed axis . . . 402
Sack, R. A.: Relaxation processes and inertial effects: II—Free rotation in space . . . 414

PAGE

Saxena, S. C., with Srivastava, B. N. : Thermal conductivity of binary and ternary rare gas mixtures . . . 369
Sayers, J., with Yeung, T. H. Y. : An r.f. probe technique for the measurement of plasma electron concentrations in the presence of negative ions . . . 663
Scaife, B. K. P. : Dipole moment fluctuations of a dielectric body . . . 314
Seymour, E. F. W., with Chapman, A. C., and Rhodes, P. : The effect of eddy currents on nuclear magnetic resonance in metals . . . 345
Singh, Sampooran : Penetration by high-speed metallic jets . . . 867
Skinner, N. J., and Wright, R. W. : The effect of the equatorial electrojet on the ionospheric E_s and F2 layers . . . 833
Smith, B. A., with Woolley, J. C. : Solid solution in the GaAs–InAs system (L) . 153
Smith, J. H., and Street, R. : Resistivity and magnetoresistance of Au_3Mn (R) . 1089
Smith, R. W., and Raynor, G. V. : The effect of liquid media on the grey tin⇌white tin transformation in commercially pure tin . . . 1135
Smith, S. D., with Moss, T. S., and Hawkins, T. D. F. : Absorption and dispersion of indium antimonide . . . 776
Smith, T., with Llewellyn, J. P. : The thermal expansion of iron ditelluride . 1113
Soane, C. M., with Cheesman, G. H. : The lattice constants of the inert elements (R) 700
Southgate, P. D. : Internal friction of single crystal silicon from 25°c to 1100°c (L) . 804
Spear, W. E. : Transit time measurements of charge carriers in amorphous selenium films . . . 669
Spreadborough, J., with Christian, J. W. : Stacking faults and dislocations in copper–aluminium and cobalt–nickel alloys . . . 1151
Spurr, R. T., and Newcomb, T. P. : The adhesion theory of friction . . . 98
Spurr, R. T., and Newcomb, T. P. : The variation of friction with velocity . . 198
Srivastava, B. N., and Saxena, S. C. : Thermal conductivity of binary and ternary rare gas mixtures . . . 369
Srivastava, G. P., with Krishnaji : Temperature variation of microwave absorption coefficient in ethyl chloride (R) . . . 621
Srivastava, R. C. : Clusius and Dickel column constants for a gas consisting of hard spherical molecules . . . 93
Starkiewicz, J., Bate, G., Bennett, H., and Hilsum, C. : A single crystal photodiode of lead sulphide (L) . . . 258
Stevens, K. W. H. : The Ising model and ferrimagnetism (L) . . . 537
Stevens, K. .W H., with Danielian, A. : High temperature susceptibilities (R) . . 326
Stokes, A. R. : The theory of light scattering by suspensions of randomly oriented long prisms . . . 379
Storey, J. R., with Farley, F. J. M. : Time variations of extensive air showers . . 840
Story, H. S., and Hoffman, R. W. : Stress annealing in vacuum deposited copper films . . . 950
Street, R., with Smith, J. H. : Resistivity and magnetoresistance of Au_3Mn (R) . 1089
Surma, M., with Piekara, A., Malecki, J., and Gibalewicz, J. : Note on production of strong magnetic fields of short duration and measurement of their intensity (R) 432
Swanson, K. M., with Andrew, E. R. : A method for determining the nculear relaxation mechanism in crystals (R) . . . 436
Sykes, M. F., with Domb, C. : High temperature susceptibility expansions foı the close packed hexagonal lattice (L) . . . 896

Tayler, R. J. : Hydromagnetic instabilities of an ideally conducting fluid . . . 31
Tayler, R. J. : The influence of an axial magnetic field on the stability of a constricted gas discharge . . . 1049
Teale, R. W., and Rowlands, G. : Some magnetothermal relations for ferromagnetics 1123
Teale, R. W., with Tebble, R. S. : Magnetothermal effects in nickel and cobalt . . 51
Tebble, R. S., and Teale, R. W. : Magnetothermal effects in nickel and cobalt . . 51
Temperley, H. N. V.: The influence of boundary conditions on the Onsager–Ising partition function for the plane square lattice . . . 192
Temperley, H. N. V. : The statistical mechanics of the steady state . . . 577
Temperley, H. N. V. : Would a classical gas of non-attracting rigid spheres show a phase transition ? (L) . . . 536

PAGE

Thompson, W. B. : Thermonuclear reaction rates 1

Tolansky, S., and Howes, V. R. : Induction of ring cracks on diamond surfaces 521

Trehan, Y. N., with Goswami, A. : The thermal decomposition of cupric oxide *in vacuo* (R) 1005

Ubbelohde, A. R., with Lewis, F. A., and Orr, J. : Contact effects resulting from compression and flash sintering in graphite powders 928

Wallace, A. A., with Cooper, R., and Grossart, D. T. : Directional effect in the electric breakdown of single crystals of KCl and NaCl 169

Walton, A. K., with Jackson, R. B. : Abnormal noise in junction transistors during secondary ionization (R) 251

Ward, L. : The temperature coefficient of reflectivity of nickel 862

Wedepohl, P. T. : Electrical and optical properties of type IIb diamonds 177

Wedepohl, P. T., with Mitchell, E. W. J. : Magnetoresistance of a p-type semiconducting diamond (R) 527

Williams, E. H. B. : A general theory of the surface friction of solids (R) 431

Wolfendale, A. W., with Gardener, M., Kisdnasamy, S., and Rössle, E. : The neon flash tube as a detector of ionizing particles 687

Wooding, E. R. : Concentration changes in an aerosol 65

Woolley, J. C., and Smith, B. A. : Solid solution in the GaAs–InAs system (L) 153

Wright, R. W., with Skinner, N. J. : The effect of the equatorial electrojet on the ionispheric E_s and F2 layers 833

Yeung, T. H. Y., and Sayers, J. : An r.f. probe technique for the measurement of plasma electron concentrations in the presence of negative ions 663

INDEX TO REVIEWS OF BOOKS

PAGE
Aller, H. : *Gaseous nebulae.* Vol. 3 *of the International Astrophysics Series* . . 630
Anglés d'Auriac, P. : *Introduction à la méchanique des milieux continus* . . . 157
Arguimbau, L. B., and Stuart, R. D. : *Frequency modulation* 809
Appleton, Sir Edward : *Science and the nation.* 1100

Barer, R. : *Lecture notes on the use of the microscope* 1021
Batchelor, G. K. (Ed.) : *Journal of Fluid Mechanics.* Vol. 1. No. 1 634
Batchelor, G. K., and Davies, R. M. (Eds.) : *Surveys in mechanics. The G. I. Taylor 70th anniversary volume* 909
Battelle Memorial Institute : *Abstracts of the literature on semiconducting and luminescent materials and their applications.* 1954 issue. 339
Bauer, E. : *Champs de vecteurs et de tenseurs* 337
Beckenbach, E. F. (Ed.) : *Modern mathematics for the engineer* 1102
Becker, Richard : *Theorie der Wärme* 157
Bergmann, L., and Schaefer, C. : *Lehrbuch der experimental Physik.* Band III. *Optik and Atomphysik.* Teil 1, *Wellenoptik* 809
Blackwood, O., and Kelly, W. : *General Physics* (2nd Edn.) 155
Blüh, Otto, in collaboration with Elder, J. D. : *Principles and applications of physics* 156
Bodewig, E. : *Matrix calculus* 905
Booth, A. D., and Booth, K. H. V. : *Automatic digital calculators* (2nd Edn) . 716
Borgnis, F. E., and Papas, C. H. : *Randwertprobleme der Mikrowellenphysik* . . 807
Born, Max : *Physics in my generation* 541
Bowden, F. P., and Tabor, D. : *Friction and lubrication* 1101
Brillouin, L. : *Science and information theory* 638
Brillouin, L., and Parodi, M. : *Propagation des ondes dans les milieux périodiques* . 155
British Journal of Applied Physics, Supplement No. 6 : *The physics of non-destructive testing* 1179
Buchdahl, H. A. : *Optical aberration coefficients* 1013
Buck, R. C. : *Advanced calculus* 1020
Buerger, M. J. : *Elementary crystallography* 814

Campbell, R. : *Théorie générale de l'équation de Mathieu et de quelques autres équations différentielles de la mécanique* 718
Carman, P. C. : *Flow of gases through porous media* 636
Chalmers, B., and King, R. : *Progress in metal physics,* VI 1021
Collins, R. B. : *High speed photography* 1178
Combustion Institute : *Fifth symposium (international) on combustion. Combustion in engines and combustion kinetics* 906
Crank, J. : *The mathematics of diffusion* 717
Cullity, B. D. : *Elements of x-ray diffraction* 911
Curry, C. : *Wave optics—interference and diffraction* 1101

Din, F. (Ed.) : *Thermodynamic functions of gases.* Volume I—*Ammonia, carbon dioxide, carbon monoxide,* Volume II—*Air, argon, acetylene ethylene, propane* 810

Eirich, F. R. : *Rheology—theory and applications.* Vol. I 1017
Emsley, H. H. : *Aberrations of thin lenses* 719
von Engel, A. : *Ionized gases* 337
Espinasse, Margaret : *Robert Hooke* 904

Fewkes, J. H., and Yarwood, J. : *Electricity, magnetism and atomic physics.* Volume I —*Electricity and magnetism* 714
Flügge, S. (Ed.) : *Handbuch der Physik.* Band I. *Mathematical methods,* I . . 910
Flügge, S. (Ed.) : *Handbuch der Physik.* Band XIX. *Electrical conductivity,* I . 912
Flügge, S. (Ed.) : *Handbuch der Physik.* Band XXI. *Electron-emission gas discharges,* I 1014

PAGE

Flügge, S. (Ed.) : *Handbuch de Physik.* Band XXII. *Gas discharges,* II . . 811

Gibson, A. F., Aigrain, P., and Burgess, R. E. (Eds.) : *Progress in semiconductors,* I 632
Giet, A. : *Abacs or nomograms* 714
Guggenheim, E. A., and Prue, J. E. : *Physiochemical calculations* 811
Guild, J. : *The interference systems of crossed diffraction gratings* 445
Guinier, A. : *Théorie et technique de la radiocristallographie* 908
Guinier, A., and Fournet, G. : *Small-angle scattering of x-rays* (translated by C. B. Walker, followed by a bibliography by K. I. Yudowitch) 341

Hiscocks, E. S. : *Laboratory administration* 905

Imai, I. (Ed.) : *Proceedings of the International Conference of Theoretical Physics.* Kyoto and Tokyo, Sept. 1953. 156
Ivall, T. E. (Ed.) : *Electronic computers: principles and applications* 808

Jones, G. O. : *Glass* 446
Jones, G. O., Rotblat, J., and Whitrow, G. J. : *Atoms in the universe* 340

Kaufmann, A. : *Mise en équations et résolution des réseaux éléctriques en régime transitoire par la méthode tensorielle* 262
Kohlhaas, R., and Otto, H. : *Roentgenstrukturanalyse von Kristallen* 442
Kopal, Z. : *Numerical analysis* 333
Kopal Z. : *Astronomical optics and related subjects* 540

Locke, A. S. : *Principles of guided missile design.* Volume I—*Guidance* . . . 636

Marton, L. (Ed.) : *Advances in electronics and electron physics.* Vol. VII . . . 443
Mentzer, J. R. : *Scattering and diffraction of radio waves* 260
Meredith, R. : *The mechanical properties of textile fibres* 335
Michels, W. C. (Ed.) : *The international dictionary of physics and electronics* . . 1022
Montgomerie, G. A. : *Digital calculating machines* 1022
Mott, B. W. : *Micro-indentation hardness testing* 1015

Oster, G., and Pollister, A. W. (Eds.) : *Physical techniques in biological research* . 158

Padgham, C. A. : *Optics and sound* 631
Page, C. H. : *Physical mathematics* 901
Pai, Shih-I. : *Viscous flow theory.* I. *Laminar flow* 1018
Pankhurst, R. C., and Holder, D. W. : *Wind tunnel technique* 342
Parr, L. M., and Hendley, C. A. : *Laboratory glassblowing* 1019
Pearcey, T. : *Table of the Fresnel integral to six decimal places* 1100
Phillips, F. C. : *An introduction to crystallography* (2nd Edn) 715
Przibram, K. : *Irradiation colour and luminescence* (Pergamon Science Series. *Physics,* Vol. I) 635

Ramsey, N. F. : *Molecular beams* 634
Robertson, J. K. : *Radiology physics* (3rd Edn) 443

Schuler, M., and Gebelein, H. : *Five place tables of elliptic functions* 713
Schuler, M., and Gebelein, H. : *Eight and nine tables of elliptic functions* . . . 713
Seitz, F., and Turnbull, D. (Eds) : *Solid state physics.* Vol. 1 334
Seitz, F., and Turnbull, D. (Eds.) : *Solid state physics.* Vol. 2 813
Sproull, R. L. : *Modern physics. A textbook for engineers* 633
Stacey, K. A. : *Light scattering in physical chemistry* 807
Stratton, J. A., Morse, P. M., Chu, L. J., Little, J. D. C., and Corbató, F. J. : *Spheroidal wave functions* 1016

PAGE

Surugue, J. : *Techniques générales du laboratoire de physique* (2nd Edn) . . . 338
Swainger, K. : *Analysis of deformation.* Vol. III (*Fluidity*) 1101

Tegart, W. J. McG. : *The electrolytic and chemical polishing of metals in research and industry* 343
Thompson, A. C., and Rodgers, O. E. : *Thermal power from nuclear reactors* . 713
Topping, J. : *Errors of observation and their treatment* 261
Townsend, A. A. : *The structure of turbulent shear flow* 542

U.S. Department of Commerce, N.B.S. Applied Mathematics Series XXXVII : *Tables of functions and of zeros of functions* 1099
U.S. Department of Commerce, N.B.S. Applied Mathematics Series XLI (reissue of Mathematical Table 8) : *Tables of the error function and its derivative* . 1099

Vajda, S. : *The theory of games and linear programming* 263
Vogel, T. : *Les fonctions orthogonales dans les problèmes aux limites de la physique mathématique* 262

White, H. E. : *Descriptive college physics* 261
Wilkes, M. V. : *Automatic digital computers* 637
Wood, A. B. : *A textbook of sound* (3rd revised Edn) 158
Wool Industries Research Association : *Wool research*, Vol. II. *Physical properties of wool fibres and fabrics* 543

(Zurich) : *Proceedings of the 2nd International Congress of the International Union for the Philosophy of Science.* Zurich, 1954. 5 Vols. 712